ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Fisica

Search for $t\bar{t}$ resonances at $\sqrt{s} = 8$ TeV in the CMS experiment at LHC

Relatore:
Chiar.mo Prof. Andrea Castro

Presentata da:
Michele Michetti

Sessione II
Anno Accademico 2015/2016

Contents

1 The Standard Model and the Z' boson **5**
- 1.1 Standard Model . . . 5
  - 1.1.1 Fermions . . . 5
  - 1.1.2 Bosons . . . 7
- 1.2 The parton model . . . 7
- 1.3 The top quark . . . 8
- 1.4 Beyond Standard Model . . . 10
  - 1.4.1 The idea for new physics . . . 11
- 1.5 Search of resonances . . . 13
- 1.6 Cross section . . . 15

2 LHC and the CMS detector **17**
- 2.1 The Large Hadron Collider . . . 17
  - 2.1.1 Creation of the pp beam at LHC . . . 18
- 2.2 The CMS detector . . . 18
  - 2.2.1 The inner tracker . . . 20
  - 2.2.2 The electromagnetic calorimeter . . . 22
  - 2.2.3 The hadron calorimeter . . . 22
  - 2.2.4 The superconducting solenoid . . . 23
  - 2.2.5 The muon spectrometer . . . 23
- 2.3 The trigger system . . . 24
- 2.4 Monte Carlo simulations . . . 24

3 $t\bar{t}$ event selection **25**
- 3.1 The event selection . . . 26
- 3.2 Definition of the objects for the analysis: the jets . . . 27
- 3.3 Relevant kinematic quantities . . . 27
  - 3.3.1 b-tag selection . . . 28
  - 3.3.2 ΔR_{bb} . . . 29
- 3.4 Event selection . . . 32
- 3.5 Background events for $t\bar{t}$ . . . 33

4 Kinematic reconstruction of the events **35**
4.1 Likelihood fit 36
4.1.1 Likelihood fit in the analysis 41
4.2 Systematic uncertainties 46
4.2.1 Jet energy scale 47
4.2.2 Integrated luminosity 47
4.2.3 b-tagging 47
4.3 Cross section upper limits 47
4.3.1 The package "theta" 48
4.3.2 Statistical model 48
4.3.3 The Theta analysis 49

Bibliography **55**

Abstract

This work concerns the search for the production of heavy resonances decaying into $t\bar{t}$ pairs at a the center-of-mass energy $\sqrt{s} = 8$ TeV in the CMS experiment at LHC.
After the Higgs boson discovery, a persisting issue in the Standard Model is the squared divergencies of the Higgs mass, that generate the "hierarchy problem". This divergence is mitigated by the Supersimmetry or, in other models, introducing the Kaluza-Klein gluonic resonances or Z′ massive resonances with higher coupling with the most massive particles. For this reason, the $t\bar{t}$ decay is the channel with the highest cross section. The models predict a mass of these resonances between 700 GeV and some TeV, and production cross sections around some picobarns.
The CMS data collected at $\sqrt{s} = 8$ TeV center-of-mass energy with integrated luminosity L= 19.5 fb^{-1} represent a large sample of data on which it is possible to look for a possible $t\bar{t}$ resonances.
In this thesis, the search is performed in the all-jets channel with reference to the resolved event topology characterised by 6 or more jets in the final state.

Abstract

Questo studio riguarda la ricerca della produzione di risonanze pesanti che decadono in coppie $t\bar{t}$ per collisioni con energia del centro di massa $\sqrt{s} = 8$ TeV con l'esperimento CMS a LHC.
Dopo la scoperta del bosone di Higgs, un problema persistente del Modello Standard è la divergenza quadratica della massa dell'Higgs, che genera il cosiddetto "problema della gerarchia". Tale divergenza è mitigata nella Supersimmetria o, in altri modelli, introducendo le risonanze gluoniche di Kaluza-Klein o le risonanze massive Z$'$, che hanno accoppiamento maggiore con le particelle più massive. Per questa ragione, il decadimento $t\bar{t}$ è il canale con la più alta sezione d'urto. I modelli predicono una massa per queste risonanze compresa fra 700 GeV a alcuni TeV a una sezione d'urto di produzione dell'ordine di alcuni picobarn.
Attualmente, i dati di CMS a energia del centro di massa di $\sqrt{s} = 8$ TeV con luminosità integrata L= 19.5 fb^{-1} sono un grande campione di dati, su cui è possibile cercare informazioni su una possibile risonanza $t\bar{t}$.
In questa tesi, l'analisi è stata eseguita nel canale all-jets con riferimento ai jet della topologia *resolved*, caratterizzati da 6 o più jet nello stato finale.

Introduction

In Standard Model (SM) and Beyond SM (BSM) searches, the top quark is fundamental for its large coupling with the Higgs boson and with the massive resonances predicted from models BSM.
The measurement of $t\bar{t}$ pairs invariant mass and cross section to compare them with Monte Carlo (MC) simulations of Quantum Chromodynamics (QDC) predictions is a powerful instrument to demonstrate the agreement between the experimental measurements and the theory.
The LHC has a large $t\bar{t}$ production cross section, allowing to record a large event data set that makes possible to study the top quark properties with a good precision.
This work searches for $t\bar{t}$ resonances in the all-jets channel using resolved events at $\sqrt{s} = 8$ TeV with integrated luminosity L = 19.5 fb^{-1}. In this decay channel both the W bosons decay into a quark-antiquark pair, so the final state of the $t\bar{t}$ events is characterized by 6 jets, with at least two of them coming from bottom quarks. The choice of this channel guarantees a high branching ratio and, with an appropriate event selection, properly discriminate the background, giving a good signal/background ratio.
The purpose of the analysis is to search for a peak in the invariant mass of the $t\bar{t}$ corresponding to the Z$'$ resonance, predicted in some models BSM.
The first chapter presents the theoretical view on which our analysis is based.
The second chapter describes the hardware and software components of the CMS detector involved in the search.
The third chapter writes about the kinematic variables used in the χ^2 statistics to select the $t\bar{t}$ events and how the MC events are used to distinguish signal and background in the data. Furthermore, the method to extract an upper limit to the Z$'$ production cross section is described.
After all, the fourth chapter show the measurements of the mass of the $t\bar{t}$ resonancies, comparing them to the theoretical predictions from MC simulations with different Z$'$ masses, and finally gives the Z$'$ production cross section upper limit results.

Chapter 1

The Standard Model and the Z′ boson

1.1 Standard Model

The Standard Model is the theory that classifies the fundamental particles and describe the strong, weak and electromagnetic interaction in the Quantum Field Theory (QFT) frame, with the symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$. Its developing was completed in the mid-1970s, upon experimental confirmation of the existence of quarks, and all of its predictions were experimentally measured as proof of its validity. The experimental confirmation of the SM prediction culminated with the Higgs boson discovery in 2012.
The SM contains 12 **fermions**, with the corresponding antiparticles, and 13 **bosons**.

1.1.1 Fermions

The fermions are particles with half-integer spin (1/2), so they follow the Fermi-Dirac statistics and respect the Pauli exclusion principle. In the SM, the fermions compose the matter and they are classified into two groups:

- **leptons**: 6 particles, of which 3 massive and with electric charge 1[1], and 3 massless and neutral, that are subjects to the electromagnetic and weak forces. They are organized in doublets:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix} .$$

[1] expressed in unity of the electron electric charge $Q_{e^-} = 1.6 \cdot 10^{-19}$ C

The row below in the doublet contains the *massive charged leptons* (electron, muon and tau), that carry electric charge and weak isospin, so they interact electromagnetically and weakly. The row above contains the *neutrinos* that carry only the weak isospin and are subject to the weak interactions. The neutrinos, in the SM, have only the left-handed component of the isospin; this mean that neutrinos are massless in the SM.
The leptons are organized in doublets because they conserve the partial and total *leptonic quantum number* when they interact.

- 6 **quarks**: massive particles that carry electric charge, weak isospin charge and colour charge, so they're subject, respectively, to the electromagnetic, weak and strong interactions. The quarks are free during an average lifetime $\tau_{had} \sim 10^{-23}$ s called *hadronization time*; at longer times, they can exists in bounded states named *hadrons*; they are distinguished in **mesons** ($q\bar{q}$) and **barions** (qqq or $\overline{qqq}$). Also the quarks are organized in doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} .$$

All the quarks and leptons quantum numbers are shown in Table 1.1.

Particle	Mass	Q_e	Y_w	C_s	Spin
e^-	0.511 MeV	-1	1/2	0	1/2
μ^-	105.7 MeV	-1	1/2	0	1/2
τ^-	1.777 GeV	-1	1/2	0	1/2
ν_{e^-}	< 2.2 eV	0	1/2	0	1/2
ν_{μ^-}	< 0.17 MeV	0	1/2	0	1/2
ν_{τ^-}	< 15.5 MeV	0	1/2	0	1/2
u	2.4 MeV	2/3	1/2	r/b/g	1/2
d	4.8 MeV	-1/3	1/2	r/b/g	1/2
c	1.27 GeV	2/3	1/2	r/b/g	1/2
s	104 MeV	-1/3	1/2	r/b/g	1/2
t	172.5 GeV	2/3	1/2	r/b/g	1/2
b	4.2 GeV	-1/3	1/2	r/b/g	1/2

Table 1.1: Quantum numbers for the SM fermions

1.1.2 Bosons

The bosons are particles with integer spin (in the SM the bosons spin is 1) that follow the Bose-Einstein statistics and are the mediators of the SM interactions; they are:

- the $\mathbf{W}^{\pm}$ and the $\mathbf{Z}^0$, mediators of the weak force. They are heavy massive particle ($m_W = 80.4$ GeV , $m_Z = 90$ GeV), in fact the weak coupling has short range ($\approx 2.5 \cdot 10^{-3}$fm). The $W^{\pm}$ has electric charge ± 1, while the Z^0 is neutral;
- 8 **gluons**, mediators of the strong coupling described in QCD. They are massless and electrically neutral, but carry a colour-anticolour charge;
- the **photon**, massless mediator of the electromagnetic coupling, as predicted in the Quantum ElectroDynamics (QED). It does not carry any colour charge and it is electrically neutral.

Then it is necessary to add the **Higgs boson**, identified at LHC by the ATLAS and CMS experiments in 2012. It is a heavy massive particle, with $m_H = 125$ GeV, spin 1 while all the other quantum numbers are zero. This means that the Higgs boson is not affected by strong, weak or electromagnetic interactions and is produced only from high-order processes. The Higgs boson gives a mass to all the SM particles.

1.2 The parton model

At LHC proton beams are constituted by many bunches with 10^11 protons for any bunch. The *pp* center-of-mass energy is a Lorentz invariant and is defined as:

$$\sqrt{s} = \sqrt{P_1^2 + P_2^2} \simeq \sqrt{2P_1P_2} = 8 \text{ TeV}, \tag{1.1}$$

where P_1 and P_2 are the 4-momenta of the colliding protons.
The protons are composed by 3 *valence quarks* (uud) and quark-antiquark pairs and gluons that are created and annihilated continuously; these components are called **partons**. So, the real centre-of-mass energy avaliable is the fraction carried from the two partons that are colliding. The i-th parton has a 4-momentum given by

$$p_i = P \cdot x_i \tag{1.2}$$

where P is the 4-momentum of the proton and x_i is the Bjorken variable. So, the centre-of-mass energy available in the collision between the proton 1

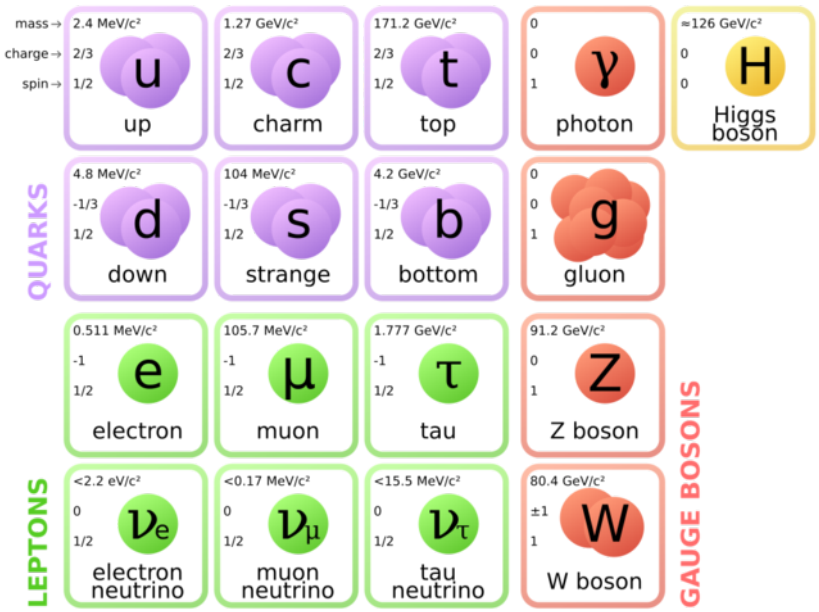


Figure 1.1: The fundamental particles of the Standard Model.

and 2 is

$$\hat{s} = (p_1 + p_2)^2 = x_1^2 P_1^2 + x_2^2 P_2^2 + 2x_1 x_2 P_1 P_2 \simeq x_1 x_2 s \tag{1.3}$$
$$\Longrightarrow \sqrt{\hat{s}} = \sqrt{x_1 x_2 s}.$$

The internal structure of the protons, described with the *parton distribution functions* (pdf), makes the event kinematics very complicated because the avaliable energy is different for each event.

1.3 The top quark

The top quark was discovered in 1995 from the experiments CDF and DØ at Fermilab (Chicago). It is the heaviest particle of the SM, with $m_t = 172.5$ GeV[2].
It is the only quark that cannot be found in a bound state with other quarks because it decays before hadronizing:

$$\tau^t_{decay} \sim 10^{-25} s << \tau_{had} \sim 10^{-23} s,$$

and it almost always decays into a W boson and a b quark; the branching ratio of this process is $\mathrm{BR}(t \to W^+ b) = 99.3\%$.
The last reason for which the top quark is important in particle physics is because it is the particle to the highest coupling with the Higgs boson.

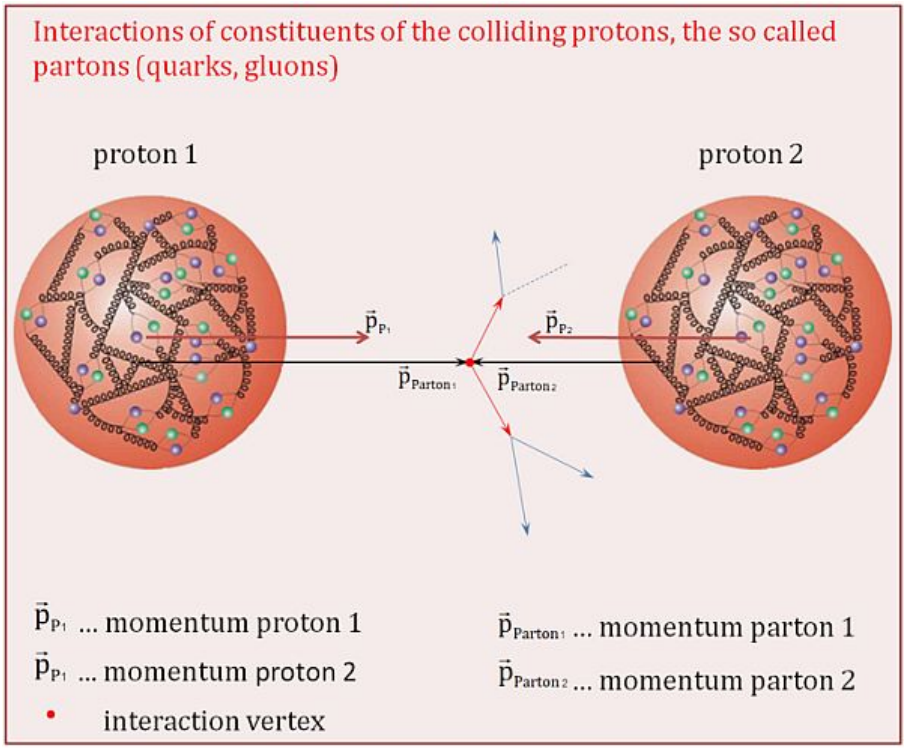


Figure 1.2: Representation of the *pp* collision in the parton model.

1.4 Beyond Standard Model

At energy scales higher than 10^{15} GeV, the SM cannot describe all the phenomena in particle physics; the new effects to consider are:

- the hierarchy problem, that consists of the divergencies of the Higgs mass at energies $> 10^{15}$ GeV;
- the mass of the neutrinos (in SM are massless), that are different for mass eigenstates and weak eigenstates; the mass hierarchy of the mass eigenstates causes the neutrino flavour oscillations, violating the conservation of the leptonic number;
- the dark matter (evaluated to be about 25% of the matter in the universe) that does not interact electromagnetically or strongly with the visible matter (5% of the mass of the universe), and the dark energy (70% of the mass of the universe) that could be the explanation of the expansion of the universe;
- the baryogenesis, that is the matter/antimatter asymmetry generated after the big bang;

To include these phenomena in a gauge theory, many BSM theories were created and they take the name of Great Unification Theories (GUT), where all the SM couplings converge to the same value at the energy scale $E_{GUT} \sim 10^{16}$ TeV. The most important ones are:

- the *M-theory* or also called *String theory*;
- the *Minimal Sequential Standard Model*;
- the Kaluza-Klein gluonic excitation or Z' massive resonances.

This work searches for massive resonances that would appear as a peak over a continuous background coming from the SM processes. In particular, this thesis studies the possible existence of Z′ massive resonances in the $t\bar{t}$ channel, with a final state characterized by 6 resolved jets, of which 2 b-tagged ones; these jets are organized in two triplets of jets, with at least one b-tagged jet for both the triplets, and both have the top quark invariant mass.

1.4.1 The idea for new physics

One of the most important question in particle physics today is whether there are new gauge bosons beyond the ones associated with the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge group.
The GUT theories are trying to unify the strong and electroweak coupling at energies $E_{GUT} \sim 10^{16}$ GeV (figure 1.3). This introduces a new spontaneous

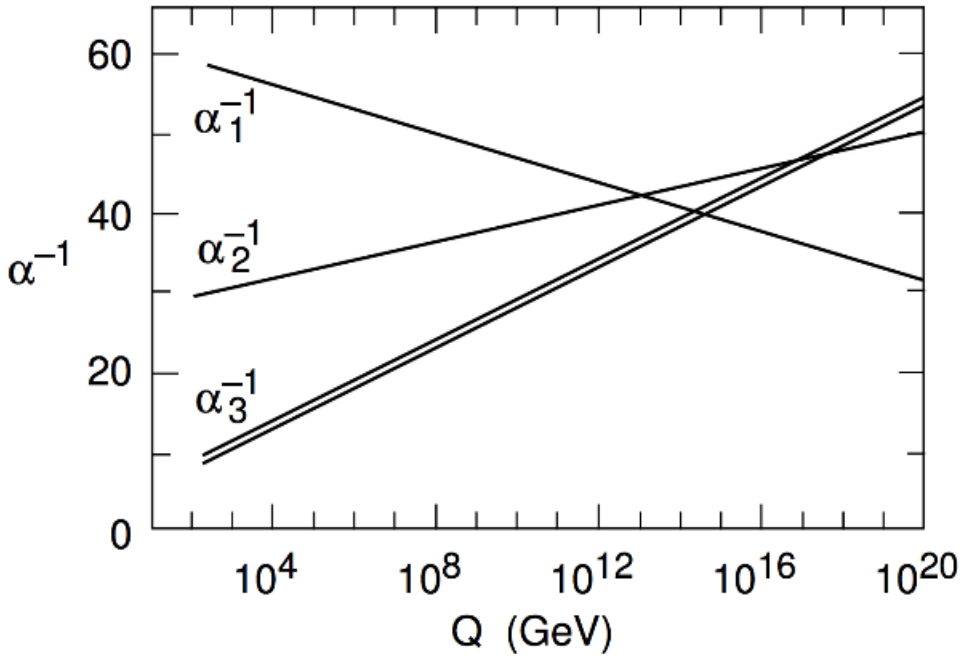


Figure 1.3: Evolution of the SM couplings $\alpha_i = \frac{g_i^2}{4\pi}$ as a function of the energy scale.

symmetry breaking (SSB) at $E << E_{GUT}$ to include the SM description at low energies that works like the SSB of the electroweak symmetry.
Some GUTs predict the spontaneous decay of the proton with the exchange of a boson with mass of the order of the Max Plank scale ($E_{Plank} \sim 10^{19}$ GeV). These theories consider the smallest group that contains all the symmetry groups: $SU(5)$. This group was abandoned because it predicted proton average lifetimes that were experimentally excluded.

Many BSM theories are built adding a new symmetry group. The easiest way is to introduce the symmetry group $U(1)'$, which corresponds to the prediction of a new Z' boson with the same Z^0 quantum number but a higher mass. The most important theories where this new heavy particle is predicted are:

- the *Left - Right Symmetric Model* (LRM);
- the *SuperSymmetry* (SUSY);
- the *String Theory*;
- the *Sequential Standard Model* (SSM);
- the *Leptophobic Topcolor Model* (LTM).

Left - Right Symmetric Model (LRM)

The new symmetry group is $SO(10)$ and it is created adding a group $U(1)_\chi$ to the symmetry group of the SM:

$$SO(10) = SU(3)_C \times SU(2)_L \times U(1)_Y \times U(1)_\chi$$

$$SO(10) = SU(3)_C \times SU(2)_L \times SU(2)_R \times U(1)_{B-L},$$

where the first chain adds to the SM a neutral gauge boson Z_χ, while the second chain introduces into the model the Left - Right Symmetry part with the groups $SU(2)_L \times SU(2)_R \times U(1)_{B-L}$ and the new neutral right-handed gauge boson Z_{LRM}.
The focus is the assignment of a right-handed component to the neutrinos, giving them a mass different from zero; so the generator of $U(1)$ becomes baryon minus lepton number (B-L), that is a physical observable. This model allows the neutrinos to have small masses.

SuperSymmetry (SUSY)

The SuperSymmetry adds to every fundamental particle of the SM a heavier particle with spin changed by 1/2 (fermions become bosons and viceversa), called the respective *supersymmetric particle*. In SUSY interactions, together with all the particles generated in the SM processes, a couple of supersymmetric particle-antiparticle is created.
These particles, if added in the loop corrections, cancel the coupling divergencies of the SM.

Superstring Theory

The string theory includes gravity into the model, describing all the fundamental particles as a one-dimensional string, instead of points with no dimensions like in the SM. Different configurations of these strings generate

all the fundamental particles that exist and also a gravitational force mediator (*graviton*). String theory requires additional space-time dimensions beyond the four SM ones.
The supersymmetric version of string theory is called Superstring Theory[8], where the gauge group introduced is E_6 which comes into play when some of the higher dimension are compactified.

Sequential Standard Model

The Sequential Standard Model (SSM) predicts a new boson Z′ with the same couplings of the Z^0 in the SM and that decays only into three fermions families. This model is not gauge invariant until the new boson has different couplings to exotic fermions or until it appear as an excited state of Z^0 boson in models with extra dimensions at weak scale[9].
So, the Z′ coupling is proportional to the square of the fermion mass with which the boson interact:

$$g_{Z'} \propto m_f^2, \tag{1.4}$$

and the decay width is

$$\Gamma_{Z'} \propto \frac{m_f^2}{M_{Z'}^2} \ . \tag{1.5}$$

This implies that $\Gamma_{Z'}$ is much larger for the decay channel $Z' \to t\bar{t}$ than for any possible decay into a fermion-antifermion couple.

Leptophobic Topcolor Model

The Leptophobic Topcolor Model[10] tries to explain the fundamental mechanism of the electroweak symmetry breaking, predicting the existence of a Z′ boson that has a strong coupling with the first and the third generation of quarks, but a small one with the leptons (this is why "leptophobic").
Like in the SSM, the Z′ decay width is proportional to the squared of the fermion mass

$$\Gamma_{Z'} \propto \frac{m_f^2}{M_{Z'}^2} \ , \tag{1.6}$$

so the decay process $Z' \to t\bar{t}$ is the favorite one.

1.5 Search of resonances

There are different processes which produce the Z′ boson and they can be divided into two categories:

- *direct*, that involve the explicit production of Z′ from fundamental particle interactions with detection of the final product from the boson decay;
- *indirect*, where we search deviations of the measured values of the SM parameters corresponding to a mixing between Z′ and Z^0 due to the extra-gauge-groups presence. This usually involves precision electroweak measurements.

The analysis we execute searches for Z′ direct production.

The SSM is taken as reference for the Z′ search due to the dependence of the Z′ coupling to the square of the fermions mass. This means that the favorite Z′ decay is

$$Z' \to t\bar{t},$$

and we have to detect the final state products of this process. From the SM, three final states are possible:

- **all-leptonic**, where both the W bosons decay into a lepton-neutrino pair:
$$Z' \to t\bar{t} \to b\bar{b}l^+l^-\nu_l\overline{\nu}_l,$$
where the leptons l can be electrons or muons and ν_l the respective neutrinos. This channel has a branching ratio BR $\simeq 9\%$ and its signature is constituted of two charged leptons, two b-tagged jets and missing energy in the transverse plane (associated to the two neutrinos);
- **semileptonic**, where a W decays into a charged lepton and a neutrino and the other into a quark-antiquark pair:
$$Z' \to t\bar{t} \to b\bar{b}q_1\overline{q}_2l^+\nu_l,$$
where $q_1\overline{q}_2$ is a quark-antiquark pair. It has a BR $\simeq 45\%$ and the signature is a charged lepton, missing transverse energy and at least two b-tagged jets;
- **all-hadronic**, where both the W bosons decay into a quark-antiquark pair:
$$Z' \to t\bar{t} \to b\bar{b}q_1\overline{q}_2q_3\overline{q}_4,$$
it has a BR $\simeq 46\%$ with a signature of no presence of charged leptons, no missing transverse energy and 6 or more jets, of which at least two are b-tagged.

Our analysis searches for the all-hadronic final state; the event selection will be discussed in chapter 3.4. All the BR for the three final states are shown in figure 1.4.

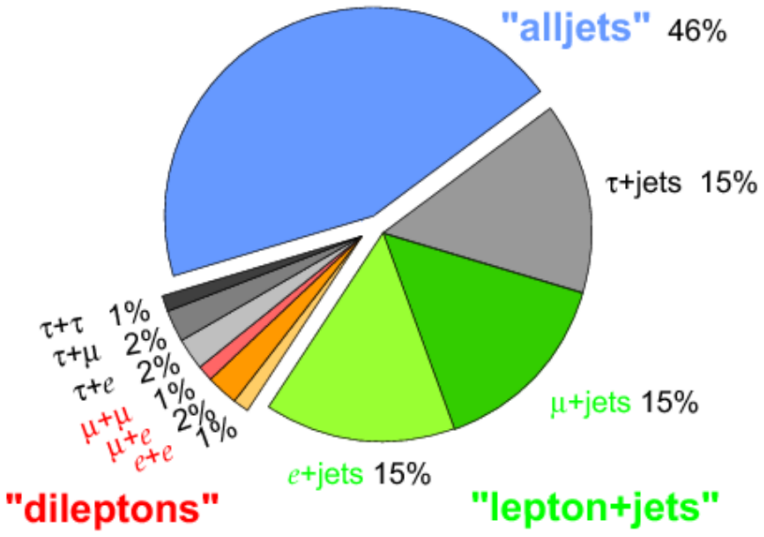


Figure 1.4: Branching Ratios for the decay channels of the $t\bar{t}$ pair.

1.6 Cross section

The cross section is a physical quantity that describes the probability to have a determined process. It is measured in cm^2 or in barn (b); the relation between them is

$$1\mathrm{cm}^2 = 10^{24}\mathrm{b} \ .$$

We are interested in the cross section $\sigma(pp \to Z' \to t\bar{t})$, that can be calculated from the BSM theories and depends on the centre-of-mass energy $\sqrt{s}$, the Z′ mass $m_{Z'}$ and the Z′ natural width $\Gamma_{Z'}$. The theoretical values at Leading Order (LO) calculated with the Leptophobic Topcolor Model[10] are summarized in table 1.2; to obtain the Next-to-Leadind Order values you have to multiply them for a factor 1.3.

From an experimental point of view, the purpose of the cross section measurements is to check if there is an agreement with the theoretical predictions. Any excess with respect to the theoretical cross section can be a hint of new physics.
Experimentally, the cross section is calculated as:

$$\sigma_{Z'} = \frac{N_{Z'}}{\epsilon \cdot L}, \tag{1.7}$$

where

- $N_{Z'}$ is the number of Z′ detected events;

$M_{Z'}$ [GeV]	$\sigma_{Z'}B(Z' \to t\bar{t})$ [pb]	
	$\Gamma_{Z'}/M_{Z'} = 1\%$	$\Gamma_{Z'}/M_{Z'} = 10\%$
500	14.79	145.06
750	3.59	33.29
1000	1.03	9.84
1250	0.367	3.37
1500	0.133	1.28

Table 1.2: Theoretical predictions for $\sigma_{Z'}B(Z' \to t\bar{t})$ calculated with the Leptophobic Topcolor Model at LO for different masses and width.

- ϵ is the Monte Carlo efficiency and is equal to

$$\epsilon = \frac{N_{MC}^{cut}}{N_{MC}^{gen}} \tag{1.8}$$

  where N_{MC}^{gen} is the number of events generated in the MC samples and N_{MC}^{cut} is the number of events that passed the selection cuts;

- L is the integrated luminosity, that is equal to 19.5 fb^{-1} in our case. It is the integral of the *instantaneous luminosity*

$$\mathcal{L} = \frac{n_1 n_2 N_b \nu}{4\pi\sigma_x\sigma_y} \tag{1.9}$$

  where n_1 and n_2 are the number of protons in the two crossing bunches, N_b is the number of bunches circulating in each direction, ν is the frequency of bunch crossing and σ_x and σ_y characterize the transverse bunch sizes.

When no signal is found we calculate an upper limit to the cross section evaluating the maximum fluctuation of the background that is still coherent with the data, as shown in chapter 4.3.

Chapter 2

LHC and the CMS detector

The physics of fundamental particles needs an adequate apparatus to be studied. For this reason at CERN, outside of Geneva, the Large Hadron Collider was built in 2007 and started to work in 2008. It has 4 major detectors: CMS, ATLAS, ALICE and LHCb.
Its construction allowed to search for SM particles into energetic regions that were not accessible before and this brought the discovery in 2012 of the last dowel of the SM: the Higgs boson. Today the general purpose experiments (ATLAS and CMS) are searching for BSM signals.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC), built in the LEP tunnel, is a circular particle accelerator that produces proton-proton (pp), Lead-Lead ($Pb - Pb$) and proton-Lead ($p - Pb$) collisions, to develop the research in particle physics.
Until 2015, LHC was colliding particles at centre-of-mass energies $\sqrt{s_max} = 7 \div 8$ TeV, now raised to 13 TeV. We analyze here the 8 TeV data, with integrated luminosity $L = 19.5$ fb^{-1}, because the rarity of the Z$'$ production process force us to use the largest samples available.
LHC is a ring made of two beam pipes and superconductor magnets, placed into a circular tunnel of 27 km of circumference at a depth between 50 m and 175 m.
Inside the two beam pipes, the beams travel in opposite directions in forced empty space at speeds near the speed of light; the beams are made of ~ 2000 "bunches" of 10^11 protons; two consecutive bunches are separated by 50 ns. The beams are mantained in circular trajectories by a 8.3 T magnetic field generated by 1232 dipole magnets, everyone of them 15 m long and cooled

with superfluid Helium at temperature 1.9 K.

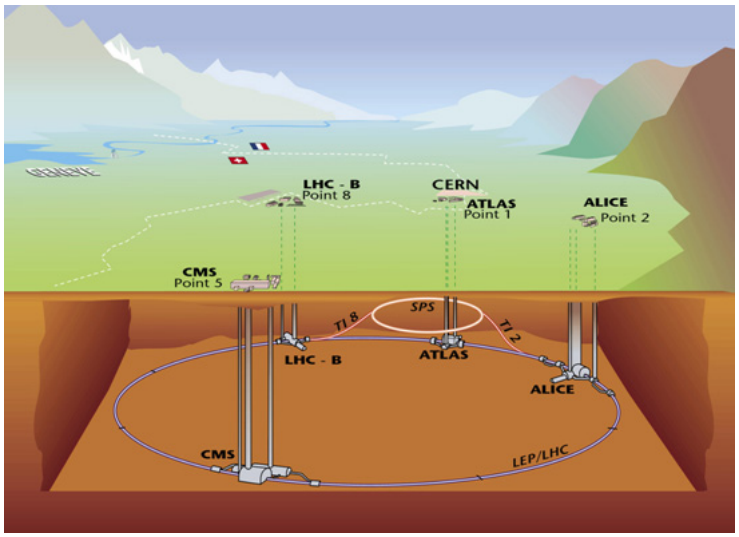


Figure 2.1: The LHC and the positions of the 4 main detectors: ATLAS, CMS, ALICE and LHCb.

2.1.1 Creation of the *pp* beam at LHC

The energy of the beam is reached in steps, through smaller accelerators that are also used today for experiments at low energy. The creation of one of the proton beam is divided in steps:

- from Hydrogen atoms, the protons are extracted removing the orbiting electrons;
- the PS Booster (PSB) accelerates the protons to 1.4 MeV;
- the beam, passed into the Super Proton Synchrotron (SPS), is accelerated to 450 GeV;
- the beams passes into the LHC and both reach the energy of 4 TeV.

2.2 The CMS detector

The Compact Muon Spectrometer (CMS) is one of the main detectors placed around the LHC ring and is general purpose. The detector is 21 m

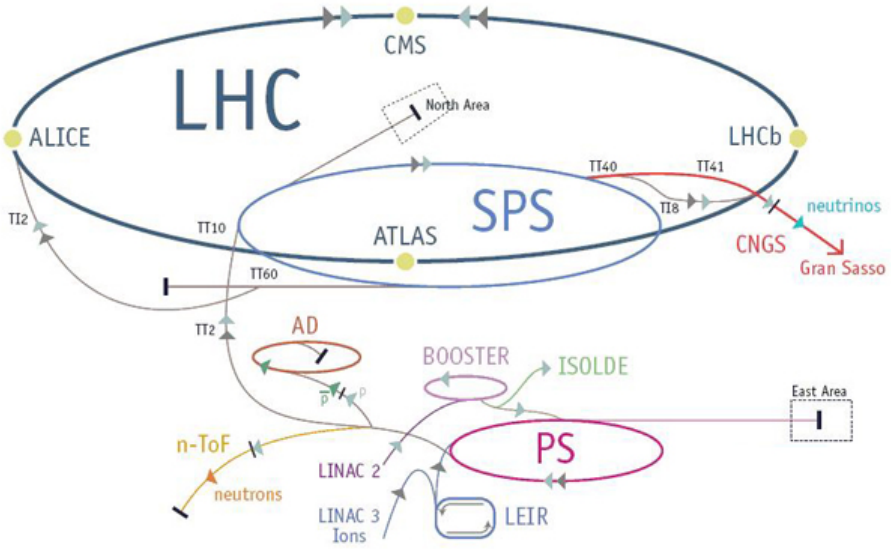


Figure 2.2: Scheme of the Large Hadron Collider and of the other acceleration rings.

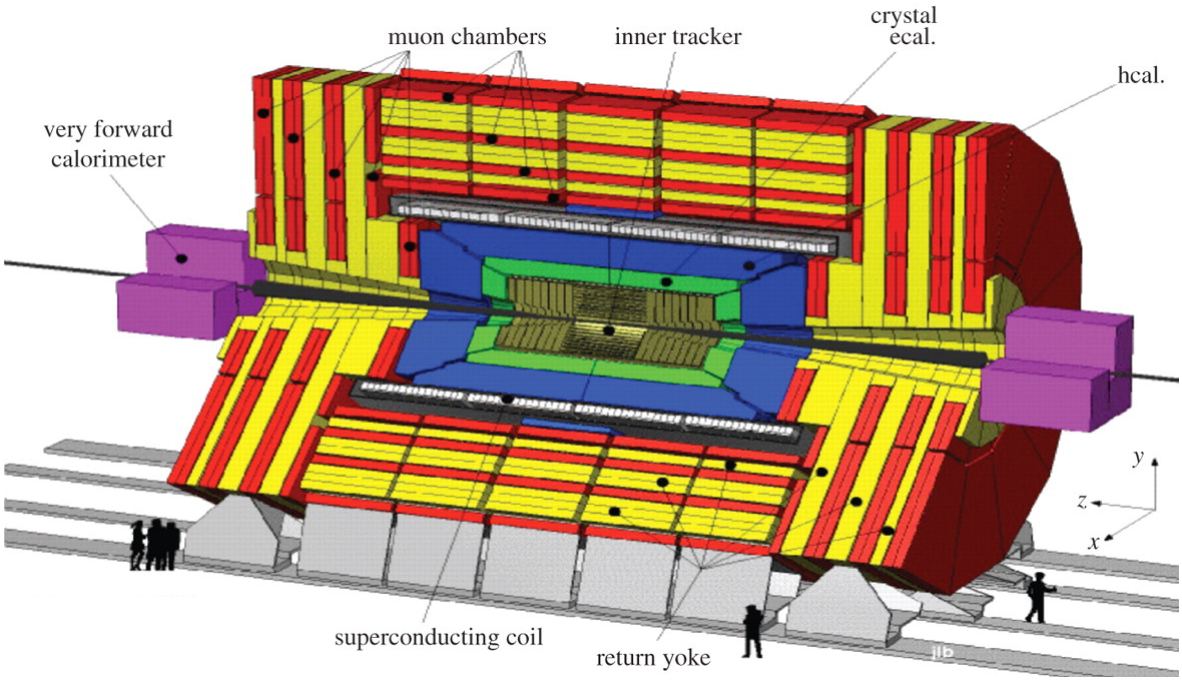


Figure 2.3: The CMS detector.

long, 15 m high and 15 m wide and weighs 12500 tonnes.
The central feature of the CMS apparatus is a 3.8 T magnetic field produced by a superconducting solenoid of 6 m internal diameter. The detector has a layered structure (see figures 2.3 and 2.4) where every layer is a different detector that has the goal to measure different quantities; from inside to outside, we can find:

- the inner tracker;
- the electromagnetic calorimeter;
- the hadronic calorimeter;
- the superconducting solenoid;
- the muon spectrometer.

2.2.1 The inner tracker

The tracker detector reveal the trajectories of the charged particles allowing to measure their linear momentum, thanks to the curved trajectory described by the interaction of the particle with the magnetic field. 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. It has to be thin to reduce the energy loss by the particle as much as it can.
The detection of many hits in the tracker determines the tracks with a resolution of ~ 10 μm.
The inner tracker measures charged particles within the range $|\eta| < 2.5$, where η indicates the detector pseudorapidity (chapter 3.3).

It is the part of the detector nearest to the pp interaction point, so it has to be radiation-resistant. It is entirely made of silicon and it has a double-layer structure:

- the first layer is the **pixel detector**. It is the one with the thinnest segmentation, because a good resolution on the track position is needed to reveal the decay products of short-lived particles of which we want to reconstruct the tracks. It contains 65 million pixels, allowing it to track the paths of particles emerging from the collision with extreme accuracy.
  Because of the huge number of channels, the power for each pixel must be kept to a minimum. Furthermore, every pixel produces 50 μwatts

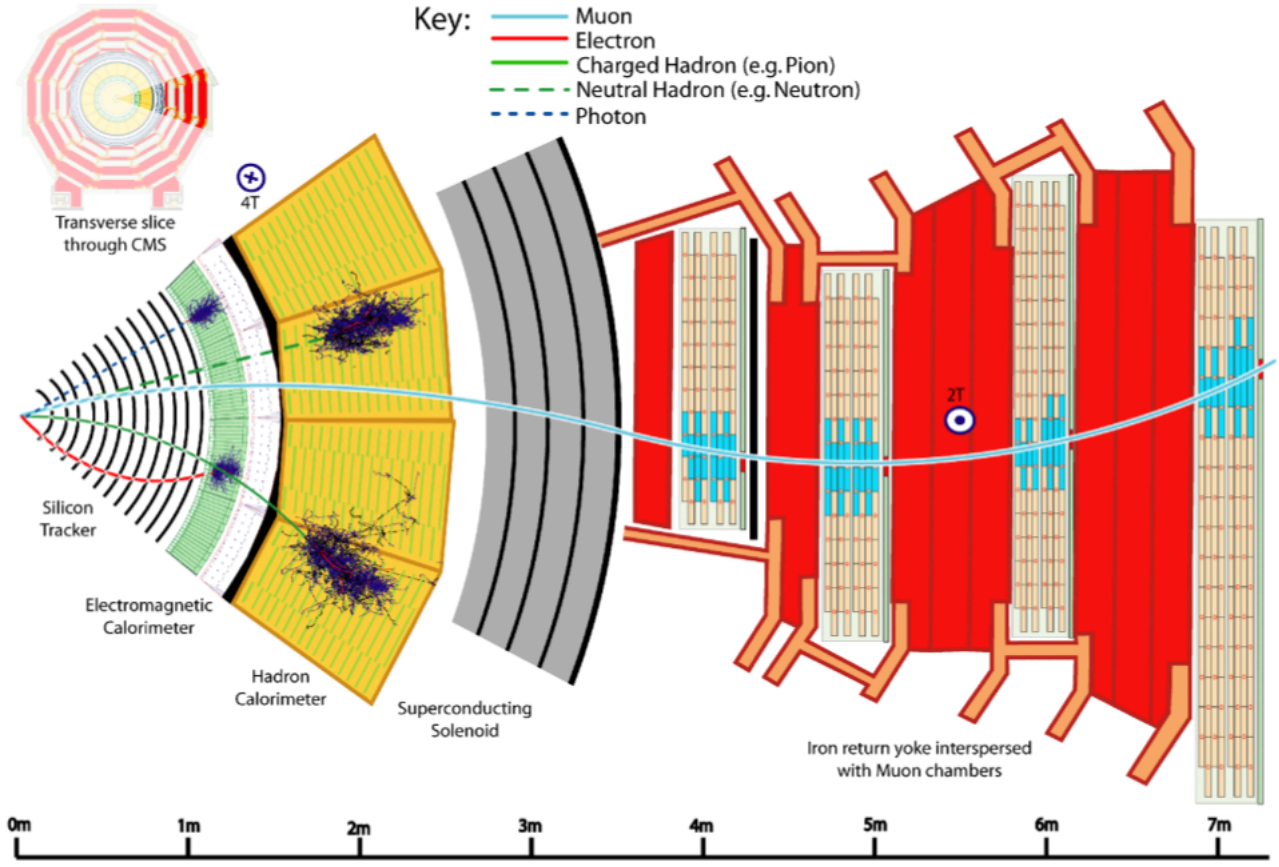


Figure 2.4: Section view of the CMS experiment. Different particles show different behaviours and trajectories within the various sub-detectors.

of energy, so the total power output is around the same as the energy produced by a hot plate. So as to not overheat the detector, the pixels are mounted on cooling tubes.

- the second layer is built with a series of strips (usually 10 strips layer). This part of the tracker contains 15200 highly sensitive modules with a total of 10 million detector strips read by 80000 microelectronic chips.

2.2.2 The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) detects the electromagnetic showers created by $e^{\pm}$ and γ. In CMS the ECAL is a *omogeneus calorimeter*, entirely made of lead tungstate crystals ($PbWO_4$) that are highly transparent scintillators. This means they produces light in proportion to the energy of the incident particles.
The ECAL , made up of a barrel section and two endcaps to coverage respectively the pseudorapidity regions $|\eta| < 1.48$ and $1.48 < |\eta| < 3.0$, is placed between the tracker and the hadronic calorimeter. It must be placed before the hadronic calorimeter because the hadronic would absorb all the e.m. shower returning a signal with fluctuations higher than the signal itself.

2.2.3 The hadron calorimeter

The Hadron Calorimeter (HCAL) measures the energy of hadrons, that are particles made of quarks, antiquarks and gluons. It also provides the indication for the presence of not detectable particles, such as neutrinos. To do this, the HCAL has to be hermetic to detect as much particles as possible from the interaction vertex. So, if we reveal interactions that does not conserve the momentum, this mean that there is missing energy carrying from invisible particles like the neutrinos.
The HCAL is a *sampling calorimeter*, so it is constituted by alternate layers of passive absorber material with high atomic number Z and density, and layers of active material that create the output signal. This choise raises from the necessity to absorb hadronic showers, that develops longitudinally and transversely much more that the electromagnetic ones. This happens because the variety of interactions and particles is larger, the nuclear interaction length λ_{int} is shorter than the e.m. radiation length X_0 and the average transverse momentum of hadronic showers is high ($< p_T >= 350$ GeV).
In the CMS detector the absorber is brass, while the active material is **fluorescent scintillator**. Special optic fibres collect up the scintillator light and

feed it into readout boxes, where photodetectors amplify the signal. These signals creates "towers" that represent the energy of the hadron passed into the HCAL.
The HCAL is divided into three sections: barrel, endcap and forward, and is the last detector before the solenoid magnet. The front-end detectors reveal all the particles that are emitted at a small angle from the beam direction, which are the most frequent particles created from the collisions; for this reason, the front and endcap calorimeters needs to be radiation-resistant

2.2.4 The superconducting solenoid

The CMS solenoid is a coil of superconducting wire that creates a magnetic field of intensity equal to 3.8T inside itself (into the tracker, ECAL and HCAL) and 2T outside (into the muon spectrometer). The superconduction condition is necessary to leave the current pass through it without resistance and generate such an intense magnetic field. The superconduction is reached cooling the magnet to 4.65 K; this is possible with a sofisticated criogenic system.
It is the largest magnet of its type ever built, so the tracker and calorimeter detectors can be placed inside the coil. This is why the detector is overall compact, compared to detectors of similar weight.
The magnet is 17 m long, with a diameter of 7 m and it weights 12000 tonnes.

2.2.5 The muon spectrometer

The main detector inside CMS is the *muon spectrometer*, how the name of the experiment suggests. Muons are leptons 200 times heavier of the electrons and they play a central role for many experiments because:

- they have a long lifetime $\tau_\mu = 2.2\ \mu$s;
- they do not interact much with the calorimeters or with the solenoid, so they conserve their initial energy;
- they are producted in the decay of many particles; for example, the Higgs boson can decay as $H \to Z^0 Z^0 \to \mu^+\mu^-\mu^+\mu^-$, that is one of its clearest signature.

For the first and second reason, the muon detectors can be placed in the most external layer of the detector.
A muon path is precisely tracked using the hits from its trajectory through the multiple layers of each of four muon station and from the inner tracker.

This gives a measurement of its momentum because we know that particles travelling with larger momentum bend less in a magnetic field. There are 1400 muon chambers, of which

- 250 are equipped with drift tubes (DTs) and 540 with cathode strip chambers (CSCs); they 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 used in the barrel region and they are displaced into 4 concentric cylinders, while for the front-end regions PRCs and CSCs are preferred.

2.3 The trigger system

When there are around 10^9 proton-proton interactions every second, the electronic cannot read out all of these events. So it is necessary to introduce 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.
There are bunch crossings every 50 ns, everyone with tens of superimposed events events, so a pipeline is needed to store the data and then retain and process informations from many interactions at the same time.

2.4 Monte Carlo simulations

The analysis is developed on data at $\sqrt{s} = 8$ TeV with integrated luminosity L = 19.5 fb^{-1} acquired in the 2015 run from the CMS experiment at LHC.
The $t\bar{t}$ signal is overhelmed by the multijet background, so the discrimination is possible using MC simulations. The MC simulations of $t\bar{t}$ events in the all-jets channel were created using POWHEG[1], that includes a next-to-leading order (NLO) QCD matrix element calculation. The detector effects are simulated using GEANT 4.
The simulation of Z′ events were carried with the MadGraph generator at Leading Order. A 100% branching ratio for $Z' \to t\bar{t}$ is assumed.
Comparing the data behavior with that of the MC simulations it is possible to notice signal and background differences in order to separate them in the clearest possible way.

Chapter 3

$t\bar{t}$ event selection

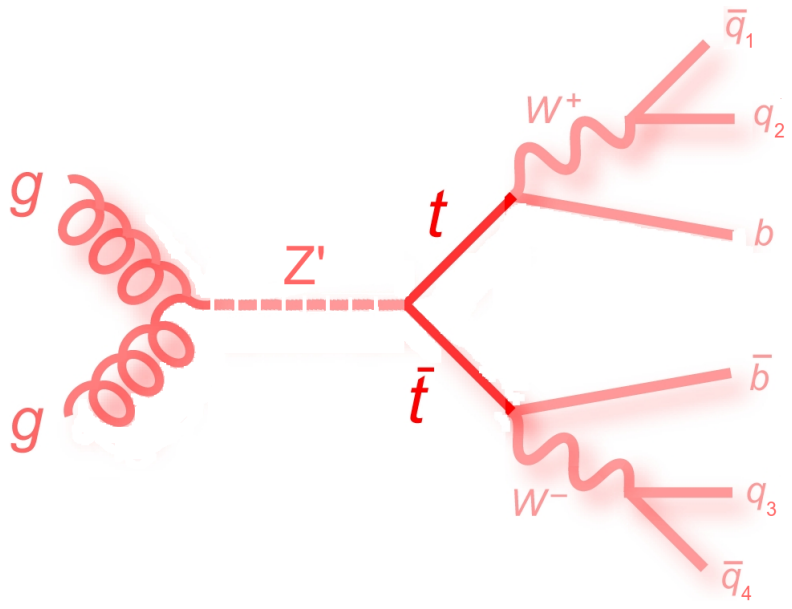


Figure 3.1: Feynman diagram for $gg \to Z' \to t\bar{t}$ process, with the all-hadronic final state.

This procedure has the purpose to choose events which come from $t\bar{t}$ pair production generated in the proton-proton collisions, rejecting as much as possible background events. From the selected events a kinematic fit then searches for the presence of a possible resonance in the invariant mass distribution.
The signal searched is associated to the all-hadronic final state (figure 3.1) of

the $t\bar{t}$ pairs, that is constituted of 6 or more jets that come from the process

$$t\bar{t} \rightarrow W^+W^-b\bar{b} \rightarrow q_1\overline{q}_2q_3\overline{q}_4b\bar{b},$$

where

- the $t(\bar{t})$ decays almost always into $W^+b(W^-\bar{b})$ (BR$\simeq$99.3%);
- the W boson from the top quark decays into $q\overline{q}'$ pair, which is composed by the combinations of quarks and antiquarks u, d, c, s, b that respect the charge of the W boson. .

The jets can be:

- of the **resolved** topology, when all the jets come from a single product of the interaction and the jets are well separated in the hadronic calorimeter;
- of the **boosted** topology, at higher energies, where the energy of the particles from the primary vertex is so high that their decay products generate jets that are so close to be unseparable because of the Lorentz boost, and are detected as a single jet.

The analysis discussed here is limited to the resolved topology jets because the invariant mass spectrum for the boosted topology jets needs a more complex algorithm to be computed. Consequently, the MC simulations used have a Z$'$ hypothesized mass between 750 and 1500 GeV because at higher mass the event reconstruction is not effective anymore.
The all-hadronic decay channel of the $t\bar{t}$ pairs has the highest branching ratio ($\sim$46%), so a large yield of multijet events is produced. However, the background events from generic QCD production (chapter 3.5) dominate the $t\bar{t}$ signal, so some techniques are used to increase the signal/background ratio (S/B).

3.1 The event selection

The selection is divided into an online phase (the trigger), that operates directly on the data that come out from the detector to reject the events useless for the analysis (low tranverse momentum, soft-gluon interactions), and an offline phase, that analyzes with macros the dataset coming from the online trigger.
The second phase is called **event selection** and rejects the events which are not interesting for our analysis with procedures that are impossible to use

online because they need a long time to be executed. The event selection uses cuts on kinematic variables as pseudorapidity η, transverse impulse p_T, angular distance between the tracks of the particles and the request for the presence of b-tagged jets.
This event selection will be followed by the kinematic fit and the invariant mass measurement.

3.2 Definition of the objects for the analysis: the jets

Jets are reconstructed offline from particle-flow candidates[3, 4] clustered by the anti-k$_T$ algorithm[5, 6] with a size parameter of 0.5. The jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be within 5 to 10% of the true momentum over the whole p_T spectrum and detector acceptance. An offset correction is applied to jet energies to take into account the contribution from additional proton-proton interactions confirmed within the same or nearby bunch crossing. Jet energy corrections are derived from simulation, and are confirmed with in situ measurements of the energy balance in dijet and photon+jet events. Additional selection criteria are applied to each event to remove spurious jet-like features originating from isolated noise patterns in certain HCAL regions. LHC produces in the 8 TeV set-up an average number of 20 interactions every proton bunch intersection; these events are called **pile-up** events. To reduce this effect, the trigger rejects the jets when the fraction of p_T of the jet tracks coming from the primary vertex of the interaction with respect to the total p_T of the tracks associated to the jet is less than 0.75. Those jets have a high fraction of tracks that do not come from the vertex of the selected events, consequently they have a high probability to be created from the pile-up collisions.

3.3 Relevant kinematic quantities

The procedure used to discriminate signal and background starts with the identification of the kinematic quantities that are indicative of the type of events that we are looking for, discriminating signal and background; the used quantities are:

- the **transverse momentum** p_T, defined as

$$p_T = p \cdot \sin\theta, \tag{3.1}$$

which is the projection of the momentum of the jet on the plane perpendicular to the beam pipe direction, and θ is the angle between the direction of the jet and the beam;

- the **pseudorapidity** η, defined as

$$\eta = -\log\left(\tan\frac{\theta}{2}\right), \tag{3.2}$$

that replaces the θ angle with an approximate Lorentz-invariant quantity;

- the angular distance between the i-jet and the j-jet defined as

$$\Delta R_{ij} = \sqrt{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}, \tag{3.3}$$

where ϕ is the azimutal angle of the jets. Detailed informations are written in the chapter 3.3.2;

- the b**-tagging discriminator** that uses a test statistic to determine if a jet is likely to come from the hadronization of a bottom quark and will be described in the chapter 3.3.1 .

All of these three variables give the possibility to perform the event selection with a series of simple cuts on them. The jets used in the analysis have preselection thresholds $p_T > 25$ GeV and $|\eta| < 2.5$.

3.3.1 b-tag selection

Jets from bottom quarks hadronization, the b-jets, are present in many physics processes like the top quark production and the Higgs boson production. For an event selection that efficiently discriminates the $t\bar{t}$ pair signal from the otherwise overhelming multijet background it is necessary to recur to a technique to determine if a jet in the final state comes from the hadronization of a bottom quark or a lighter quark.
The multijet background processes are dominated by QCD events containing jets from gluons, light quarks (u, d, s) and c-quark fragmentation.
The hadrons containing bottom quarks have relatively large masses, long lifetimes and daughter particles with hard momentum spectra and large impact parameter.
The CMS detector has a robust algorithm, based on the positions of the

interaction vertexes, the charged-particles tracking and the particles identification, that can include all the quantities to classify a jet as b-tagged or not in a discriminator variable which assumes values between 0 and 1: the higher its value, the higher is the probability to have a jet associated to a b quark. So, the discrimination is reduced to a simple cut on this quantity described in Fig 3.2; this technique is known as the **CSV b-tagging**[7], for which there are three operating points:

- *loose threshold* = 0.244 which has a high b-tag efficiency;
- *medium threshold* = 0.679 with average efficiency and purity;
- *tight threshold* = 0.898 that has a high purity.

These threshold have a probability of misidentifying a light-parton jet close to 10%, 1% and 0.1%, respectively.
To reject the high QCD background, the misidentification probability was minimized choosing the **tight** operating point. This is necessary because we need to strongly suppress the background.
The $t\bar{t}$ pair events are characterized by at least two jets from a bottom quark, so the request is:

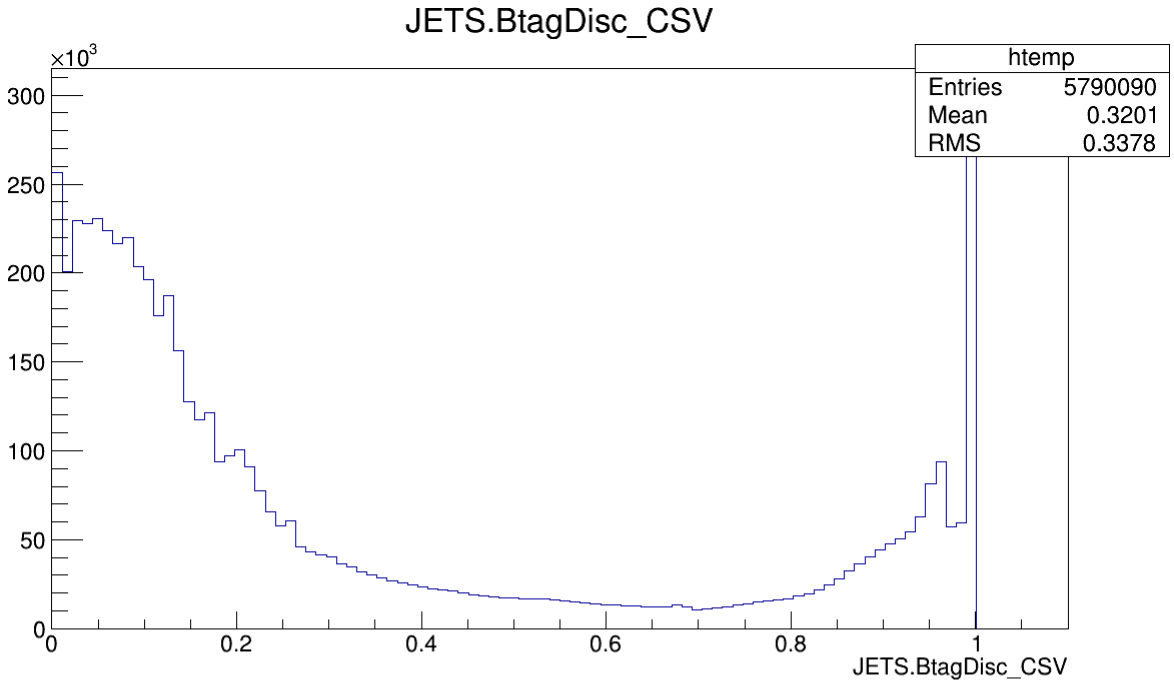
$$N_{b-tag} \geq 2.$$

3.3.2 ΔR_{bb}

The ΔR_{bb} variable, defined as the ΔR variable between two b-tagged jets, allows to reject the events that are not compatible with the emission in opposite directions of the top quark and antiquark produced in the proton-proton collisions. The cut was set comparing the behaviour of ΔR_{bb} for MC $t\bar{t}$ events and data events (dominated by QCD background), as shown in figure 3.3.

The difference between the two distributions is due to different production mechanisms of bottom quarks in background events. They can be divided into three topologies:

- *Direct production*: the two bottom quarks are produced almost back-to-back, with $\Delta R_{bb} \simeq 3$;
- *Gluon splitting*: a gluon decays into two bottom quarks that are emitted with a small angle between them, usually $\Delta R_{bb} < 1$;
- *Flavor excitation*: includes all the processes with Feynman diagrams at orders higher than LO, with a intermediate value of ΔR_{bb}.



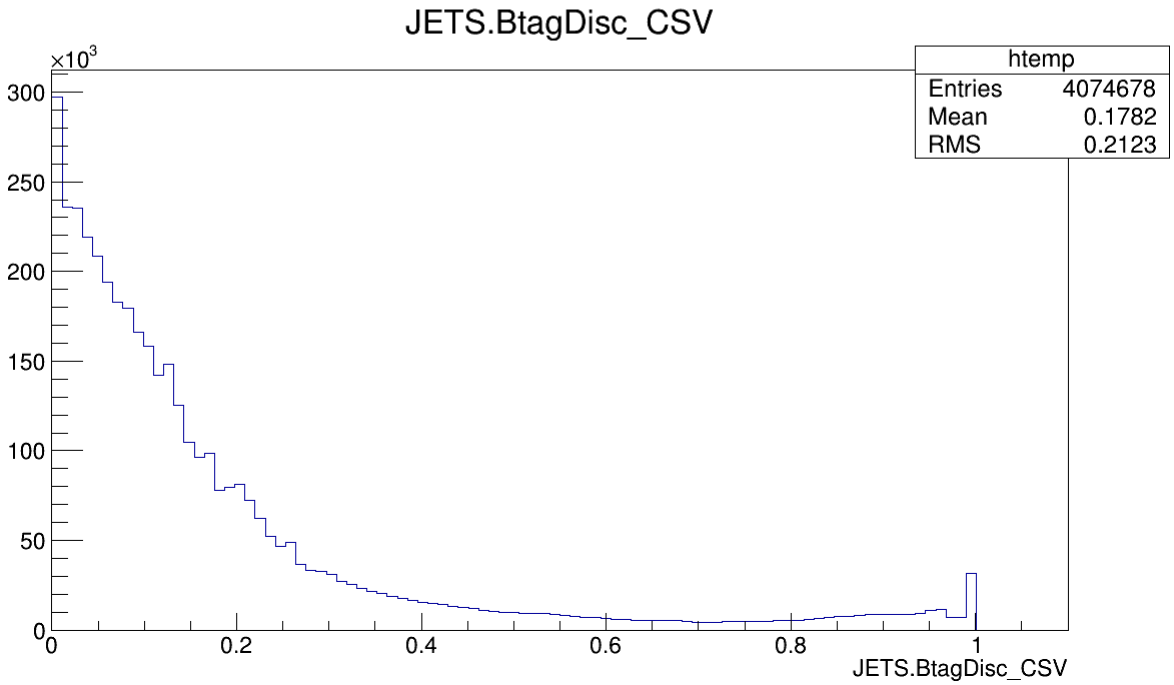


Figure 3.2: CVS b-tagging discriminator for $t\bar{t}$ MC events (a) and data, i.e. generic multijet events (b)

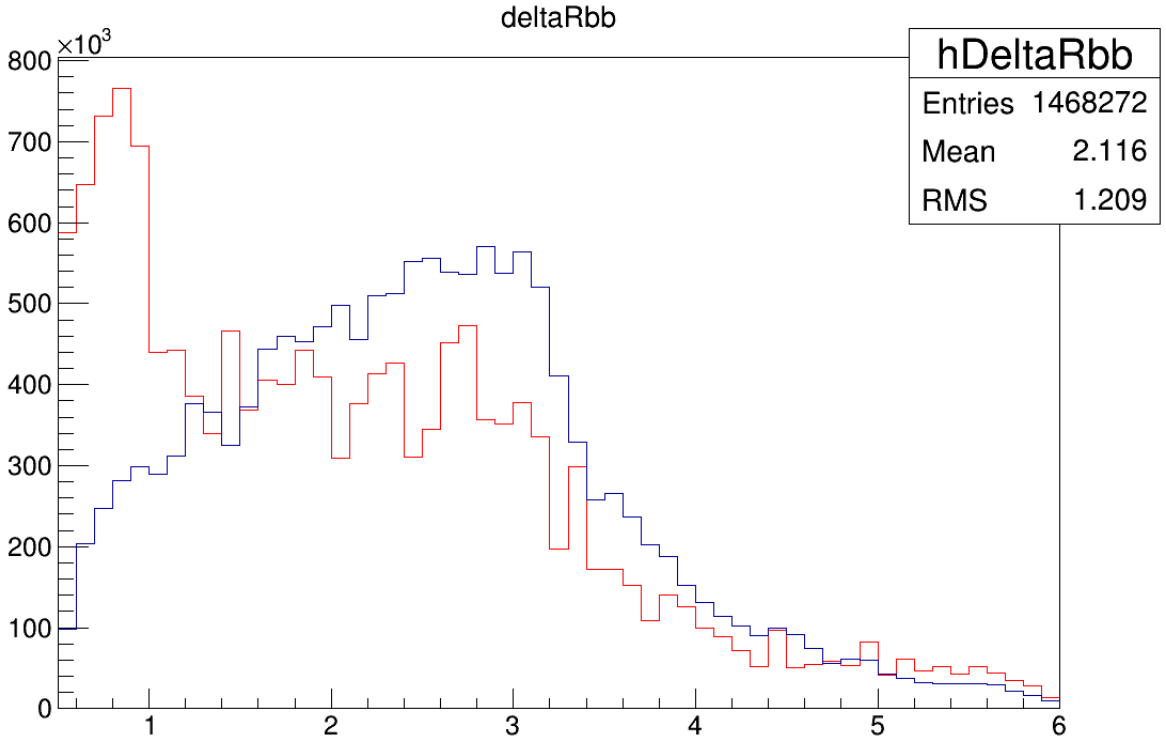


Figure 3.3: ΔR_{bb} distribution for the $t\bar{t}$ MC events (blue line) and the data (red line).

The $t\bar{t}$ signal is favored when the b-tagged jets are widely separated, so the cut on this quantity is set as

$$\Delta R_{bb} \geq 1.5,$$

which allows to reject all the gluon splitting background and part of the flavor excitation one, preserving a large fraction of signal events.

3.4 Event selection

Data are saved in the ROOT framework as TTrees structures. The selection of events which are candidates to be from $t\bar{t}$ pair production is executed with a ROOT macro, written in C++, and it is based on a series of cuts on the aforementioned kinematic variables.
The event selection starts with the requirement for events with at least 6 jets, with $p_T > 30$ GeV and $|\eta| < 2.4$ for each jet, to reject soft-interactions events and events with low p_T.
The second step accepts the events with at least 2 b-tagged jets with $|\eta| < 2.4$ and $p_T > 50$ GeV.
The third step requires (see figure 3.3):

$$\Delta R_{bb} > 1.5. \tag{3.4}$$

Since the jets will be associated to the $t\bar{t} \to W^+W^-b\bar{b} \to q_1\overline{q_2}q_3\overline{q_4}b\bar{b}$ hypothesis, the last step concerns the rejection of all the permutations where the jet association to the W and b candidates is not correct, that is:

- the W^+ from the top with the $\bar{b}$ of the antitop;
- the W^- from the antitop with b of the top.

This effect, if not removed, would result in an artificial peak in the top-antitop invariant mass distributions. The cut used to reject these events is

$$\frac{\Delta R_{W^+b} + \Delta R_{W^-\bar{b}}}{2} < 3.0 \tag{3.5}$$

where ΔR_{W^+b} and $\Delta R_{W^-\bar{b}}$ are the ΔR between the W^+b and the $W^-\bar{b}$, that are the decay products of the top and antitop respectively. The selected events then are reconstructed with a kinematic χ^2 fit (see chapter 4).

3.5 Background events for $t\bar{t}$

The major background for the events selected in this analysis comes from QCD multijets events. Additional backgrounds of lower relevance are the W+jets and single top quark production. Simulating the QCD background would be problematic in terms of CPU time which would be needed and also because of large theoretical uncertainties which are present in the cross section calculation. We avoid these problems by recurring to a procedure which infers the multijet background from the data themselves. We thus recur to the 0-tag technique: the analysis is repeated on the data with the same request, but associating the b partons to two random jets and not to b-tagged jets. The simulated $t\bar{t}$ events together with the data-based model of the background will be then used to estimate the fractions of signal and background events in the selected data.

Chapter 4

Kinematic reconstruction of the events

The kinematic fit has the goal of reconstructing the kinematics of the $t\bar{t}$ events, selecting for each event the permutation of 6 jets of the event that has the best grade of comparison to the $t\bar{t}$ hypothesis. The $t(\bar{t})$ production is associated to a jet triplet with one b-tagged jet and two jets from the W decay. Furthermore, the fit asks that the mass of two jets from the W boson is near to the W mass.The permutation selected for every event is the one that minimizes the χ^2 function

$$\chi^2 = \sum_{i=1}^{2} \left(\frac{(M_t^{reco} - M_{t,i}^{meas})^2}{\sigma_T^2} + \frac{(M_W - M_{W,i}^{meas})^2}{\sigma_W^2} \right) + \sum_{i=1}^{6} \frac{(p_{T,i}^{reco} - p_{T,i}^{meas})^2}{\sigma_p^2}, \tag{4.1}$$

where

- M_t^{reco} is the top quark mass reconstructed from the kinematic fit. The same value is used for both triplets;
- M_W is the W boson mass, fixed to 80.4 GeV[2];
- $M_{t,i}^{meas}$ is the top quark mass as measured from the invariant mass of the i-th jet triplet;
- $M_{W,i}^{meas}$ is the W boson mass measured from the invariant mass of the i-th jet doublet;
- $p_{T,i}^{reco}$ and $p_{T,i}^{meas}$ are, respectively, the recostructed transverse momentum from the kinematic fit and the measured transverse momentum of the i-th jet;

- σ_t, σ_W and σ_p are the uncertainties on the measured top quark mass, the W boson mass and the jets p_T, respectively.

The MINUIT code (with MIGRAD algorithm) minimizes the χ^2 function for every event, selecting the permutation of the 6 jets with the lowest χ^2 value and extracting the M_t^{reco} and the $p_{T,i}^{reco}$ values. The distributions of the corresponding (best) values for χ^2 and M_t^{reco} are shown in figure 4.1 and 4.2, respectively.
Then, the invariant mass of the 6-jets final state is calculated as:

$$M_{t\bar{t}}^{reco} = \sqrt{\left(\sum_{i=1}^{6} E_i\right)^2 - \left(\sum_{i=1}^{6} \overrightarrow{p_i}\right)^2} \quad , \tag{4.2}$$

where the energy E_i and the momenta $\overrightarrow{p_i}$ are those associated to the value $p_{T,i}^{reco}$ returned by the fit, assuming no change to the well-measured jet directions. Its distribution is shown in figure 4.3 for simulated $t\bar{t}$ events and for the multijet data.

4.1 Likelihood fit

The next step is to compare the $M_{t\bar{t}}^{reco}$ distributions for data, MC simulations of $t\bar{t}$ events and multijet background. In order to do this it is necessary to scale the MC signal and the background mass spectra to the corresponding expected yields. The M_t^{reco} distribution (figure 4.2) is quite distinctive, so it is used to calibrate the parameter which scales the distributions of the $t\bar{t}$ mass. The parameter is the fraction of signal events, f_s.

In statistics, a likelihood function $\mathcal{L}$ is a function of the parameters of a statistical model. It is constructed multiplying the probability density function $P_i(x_i)$ of every quantity on which the analysis is dependent:

$$\mathcal{L} = \prod_i P_i(x_i) \ . \tag{4.3}$$

The likelihood is crucial, for statistical inference, to find parameters of certain distributions from a set of statistics. For a set of parameters $\overrightarrow{\alpha}$ and a set of measurements $\overrightarrow{x}$, the $\mathcal{L}(\overrightarrow{\alpha})$ is the probability to have that set of measurements, given the variation of $\overrightarrow{\alpha}$:

$$\mathcal{L}(\overrightarrow{\alpha}) = P(\overrightarrow{x}|\overrightarrow{\alpha}) \tag{4.4}$$

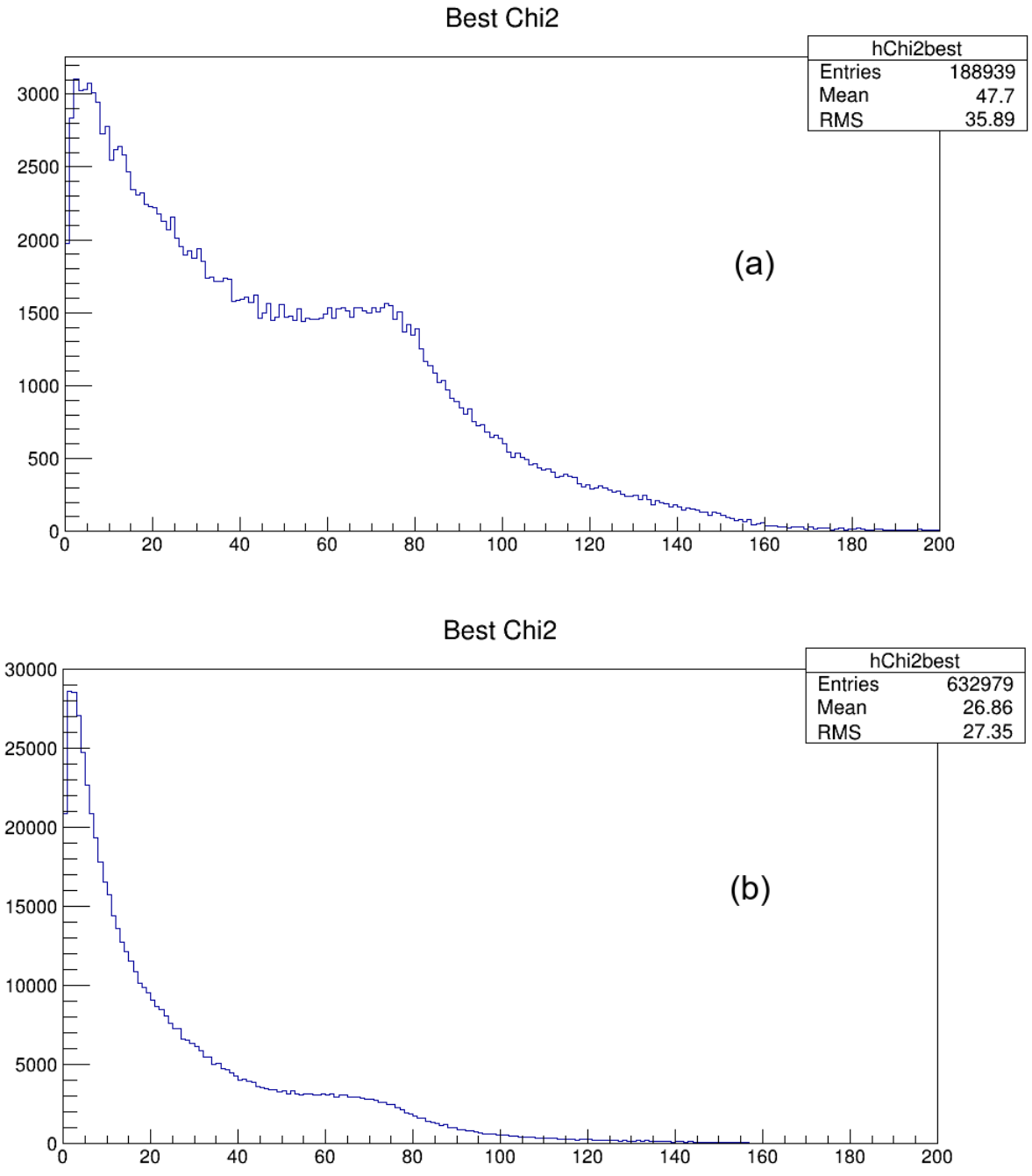


Figure 4.1: χ^2 distributions for multijet data (a) and $t\bar{t}$ MC events (b).

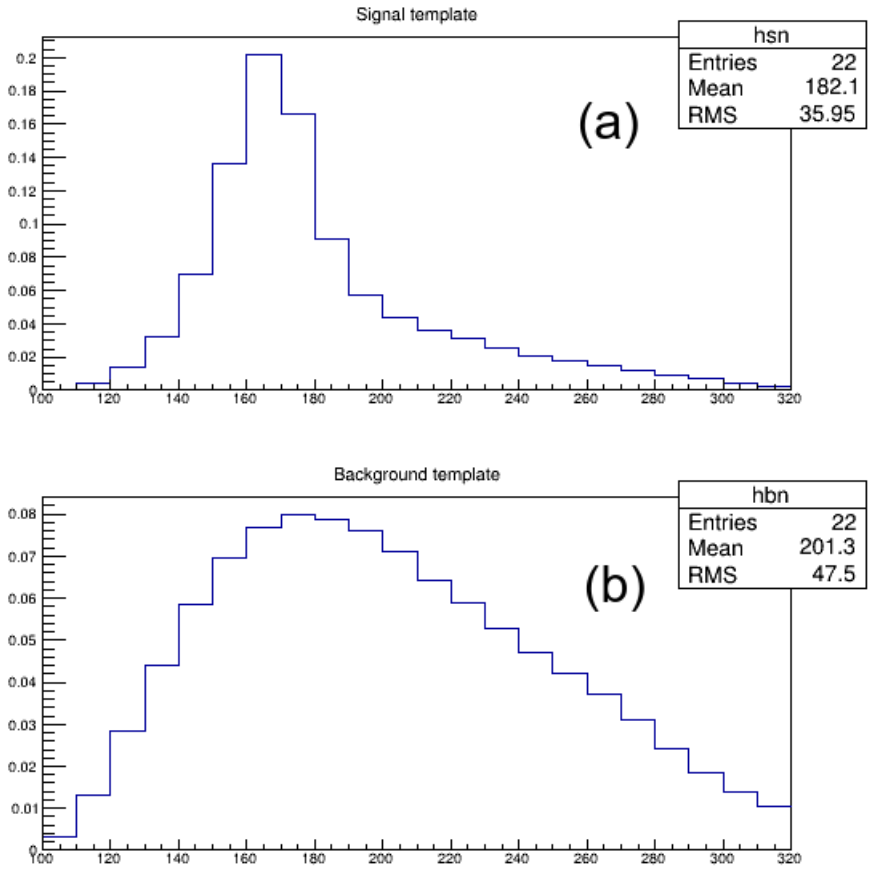


Figure 4.2: M_t^{reco} distributions for $t\bar{t}$ events (a) and the multijet background events(b), with $\chi^2 < 15$.

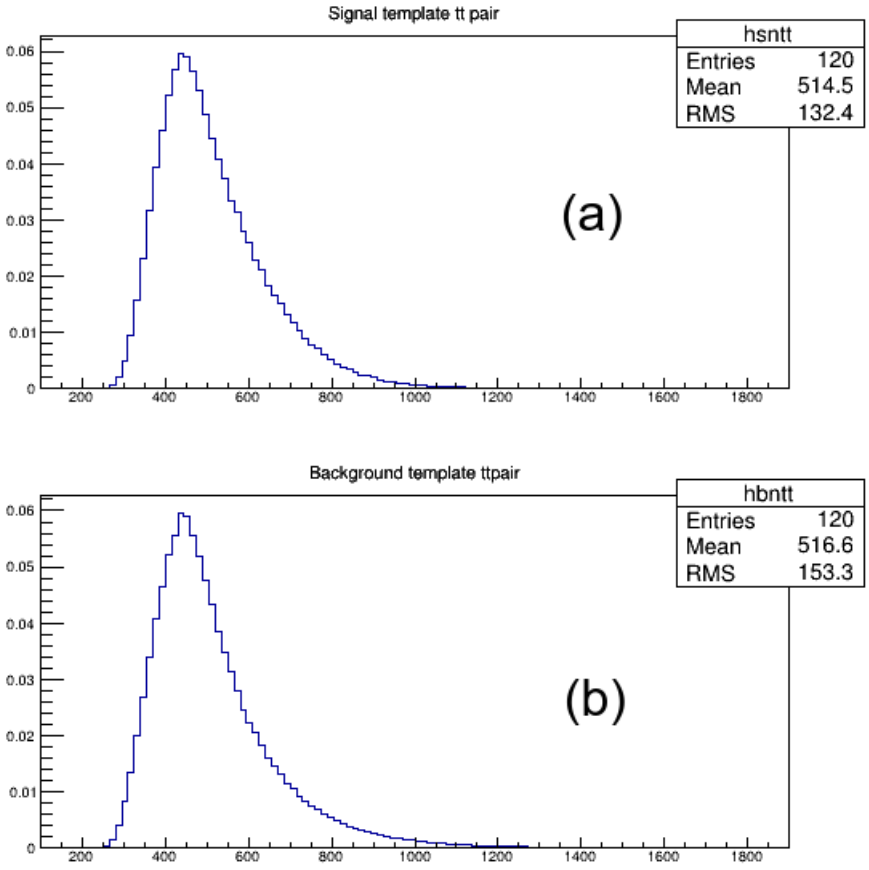


Figure 4.3: $M_{t\bar{t}}^{reco}$ distributions for $t\bar{t}$ events (a) and the multijet background events (b), with $\chi^2 < 15$.

The number of events N has a Poissonian distribution

$$P(N|\mu) = \frac{\mu^N}{N!} e^{-\mu} \ , \tag{4.5}$$

where μ is the expectation value of the number of $t\bar{t}$ and background events (n_s and n_b respectively) and its variance is $\sigma = \sqrt{N}$ .This distribution is called the *normalization likelihood* $\mathcal{L}_{norm}$.

Instead, the likelihood part that describe the population of the bins of the distribution is called the *shape likelihood* $\mathcal{L}_{shape}$ and is calculated as

$$\mathcal{L}_{shape} = \prod_{i=1}^{N_{bins}} \frac{n_s \mathcal{P}_s^i + n_b \mathcal{P}_b^i}{n_s + n_b}, \tag{4.6}$$

where

$$\mathcal{P}_s^i = \frac{n_s^i}{\sum_{i=1}^{N_{bins}} n_s^i} \qquad \mathcal{P}_b^i = \frac{n_b^i}{\sum_{i=1}^{N_{bins}} n_b^i} \tag{4.7}$$

are, respectively, the normalized distributions for $t\bar{t}$ events and multijet background events, which represent the corresponding *probability density function* (templates shown in figure 4.2). The outcome of this likelihood fit is shown in figure 4.4 for two possible cuts on χ^2. The two populations ($t\bar{t}$ events and multijet background) are well distinguished by the fit. Using f_s fraction returned by the fit to normalize the two contributions, thir sum agrees well with data.
The total likelihood is the product of the two likelihoods:

$$\mathcal{L} = \mathcal{L}_{norm} \cdot \mathcal{L}_{shape} = \frac{\mu^N}{N!} e^{-\mu} \cdot \prod_{i=1}^{N_{bins}} \frac{n_s \mathcal{P}_s^i + n_b \mathcal{P}_b^i}{n_s + n_b} \ . \tag{4.8}$$

Numerically, minimizing a function is easier than maximizing it, so the MINUIT program receives as input the negative logarithm of the likelihood:

$$-\ln \mathcal{L} = -\left\{ -\mu + N \ln \mu - \ln N! + \sum_{i=1}^{N_{bins}} \left[\ln \left(n_s \mathcal{P}_s^i + n_b \mathcal{P}_b^i \right) - \ln \left(n_s + n_b \right) \right] \right\} \tag{4.9}$$

which becomes

$$-\ln \mathcal{L} = \mu - N \ln \mu + \ln N! - \sum_{i=1}^{N_{bins}} \ln \left(\frac{n_s \mathcal{P}_s^i + n_b \mathcal{P}_b^i}{n_s + n_b} \right) =$$

$$= \mu - N \ln \mu + \ln N! - \sum_{i=1}^{N_{bins}} \ln \left[\frac{n_s \mathcal{P}_s^i + N(1 - f_s^i) \mathcal{P}_b^i}{n_s + n_b} \right] \ . \tag{4.10}$$

4.1.1 Likelihood fit in the analysis

The $-\log \mathcal{L}$ quantity is minimized by MINUIT (with the MIGRAD algorithm), returning the number of $t\bar{t}$ events n_s in the data or, altrnatively, the fraction f_s. Then, this value is used to scale the $M_{t\bar{t}}^{reco}$ distributions for $t\bar{t}$ events and multijet background according to their expected yields.
The request $\chi^2 < 5$ would provide a smaller background but is not considered further here. In fact the $M_{t\bar{t}}^{reco}$ distributions would have large fluctuations, given the current small size of the simulated Z′ samples. The ROOT macro that executes the likelihood fit return the results summarized in table 4.1.

χ^2 cut	f_s	N_{event}	n_s	n_b
$\chi^2 < 15$	0.64073	21080	13441	7639
$\chi^2 < 5$	0.674863	7718	5340	2378

Table 4.1: Fractions and number of $t\bar{t}$ and background events from χ^2 fit.

To reduce the multijet background, we require the events to have a M_t^{reco} value in the range $150 - 200$ GeV; the events passing this additional request are then used to fill the $M_{t\bar{t}}^{reco}$ distributions, scaling them according to the fraction of events expected to fall within the window 150-200 GeV.

This procedure is repeated for the MC simulations of $Z' \to t\bar{t}$ production for different Z' masses and assuming widths $\Gamma_{Z'} = 1\% M_{Z'}$ (Table 4.2) and $\Gamma_{Z'} = 10\% M_{Z'}$ (Table 4.3).
The $t\bar{t}$ mass spectrum (see figure 4.5) has the y axis drawn in logarithmic scale to show the very small $Z' \to t\bar{t}$ signal compared to the data and the sum of $t\bar{t}$ events and multijets background distributions.
In figures 4.6 and 4.7 are shown the $M_{t\bar{t}^{reco}}$ distributions for $Z' \to t\bar{t}$ MC signals for the five considered masses, with $\Gamma_{Z'} = 1\% M_{Z'}$ and $\Gamma_{Z'} = 10\% M_{Z'}$ respectively. The kinematic reconstruction for resolved jets is not effective for invariant masses higher than 1250 GeV because of the dominant presence of boosted jets over the resolved ones: the efficiency decreases and the $M_{t\bar{t}}$ peak tents to wash-out.

From the comparison between the $M_{t\bar{t}}^{reco}$ distributions we can observe that:

- the data are in good agreement with the predicted distribution for $t\bar{t}$ and multijet background for;
- there is no evidence for resonances with width as expected from narrow or wide simulated Z′ events;

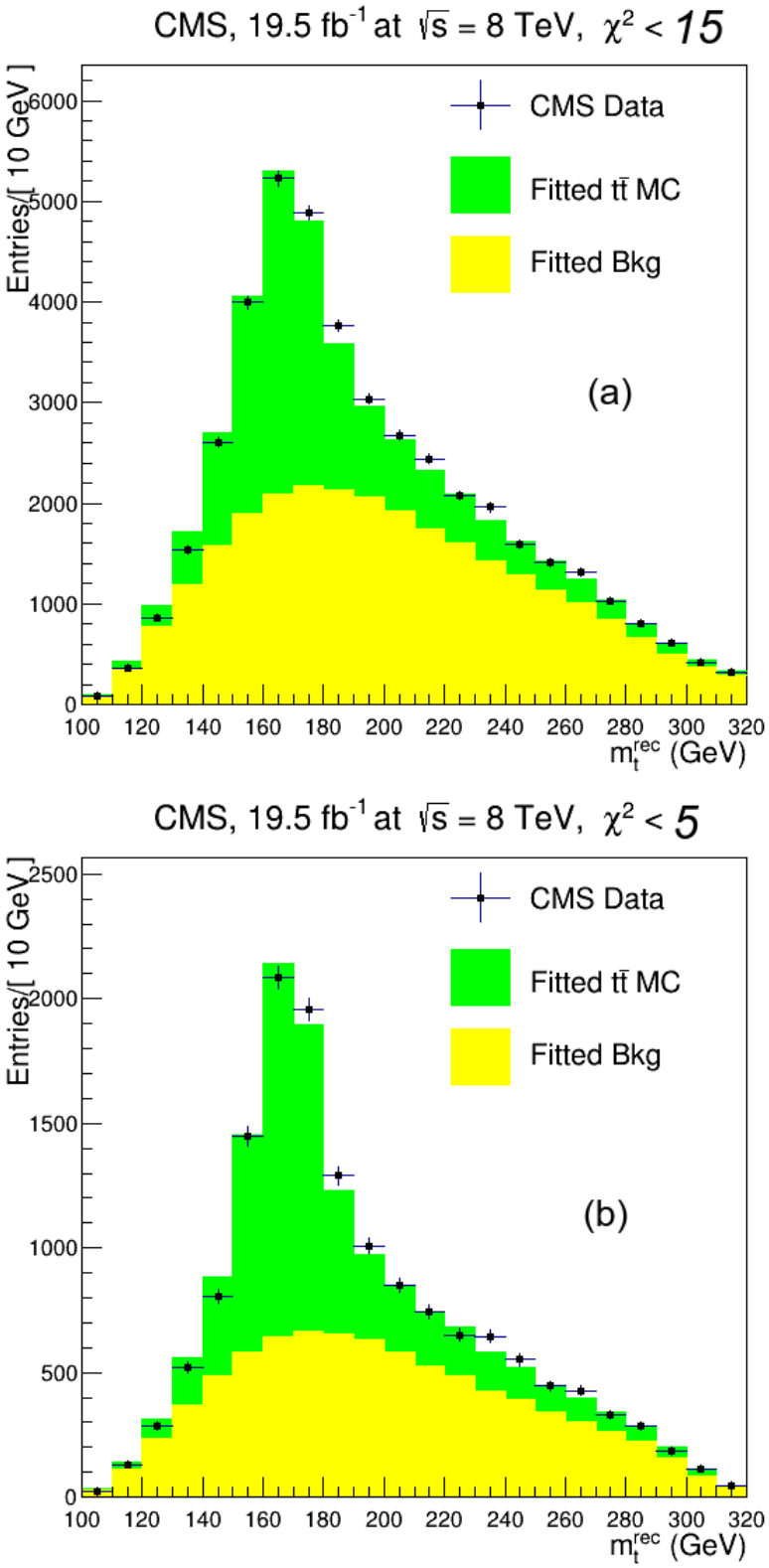


Figure 4.4: M_t^{reco} distributions for $\chi^2 < 15$ (a) and $\chi^2 < 5$ (b). The $t\bar{t}$ and background yields are normalized to the fit outcome.

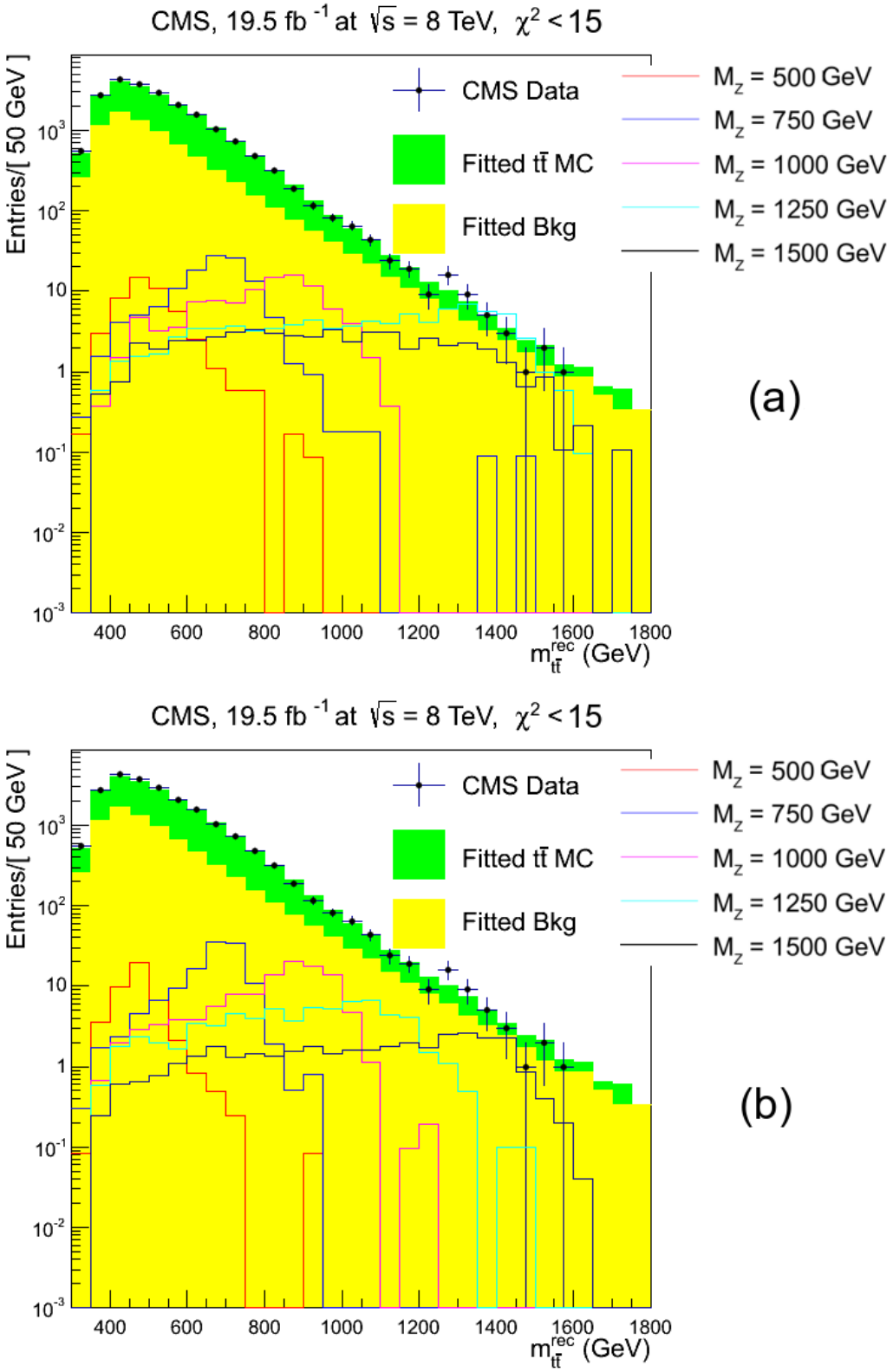


Figure 4.5: $M_{t\bar{t}}^{reco}$ for $t\bar{t}$ events and for multijet events with $\chi^2 < 15$ and 150¡M_t^{reco}¡200 GeV. Shown are MC distributions for several Z' simulations with different masses and with different width: $\Gamma_{Z'} = 10\%$ (a) and $\Gamma_{Z'} = 1\%$ (b), assuming a production cross section of 1 pb.

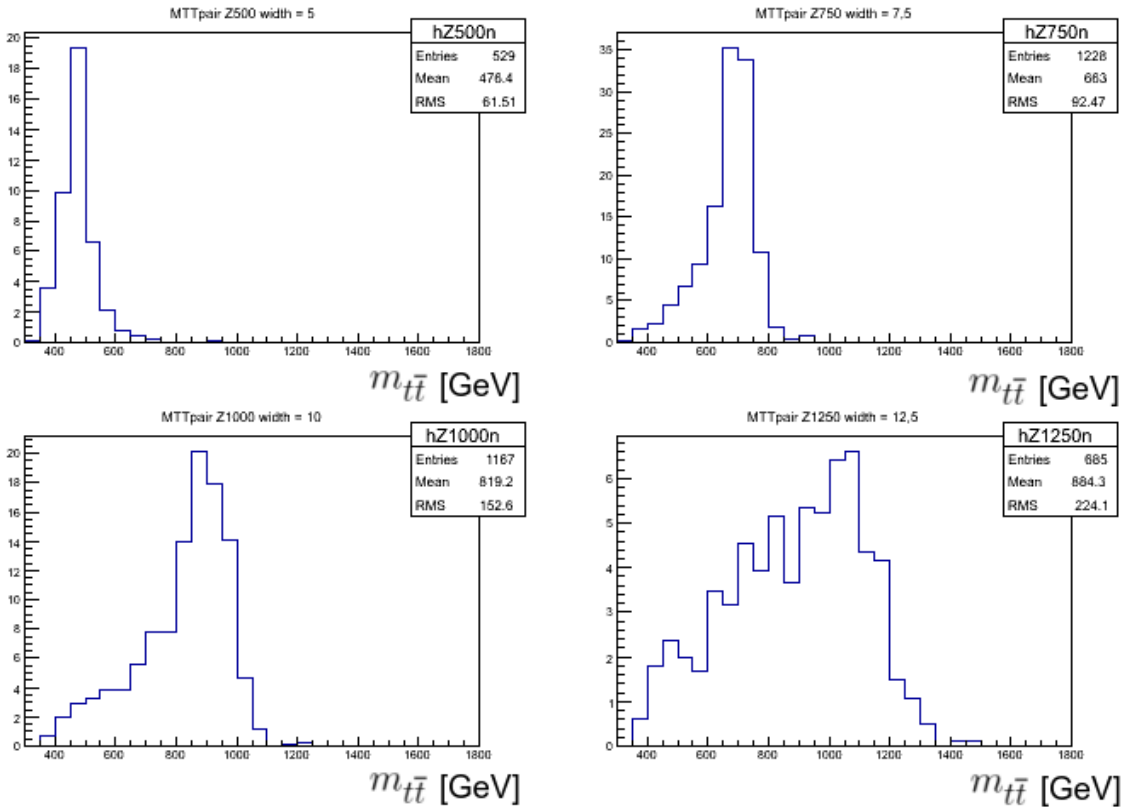


Figure 4.6: $M_{t\bar{t}}^{reco}$ distribution for Z' MC events passing the $\chi^2 < 15$ request and with 150< M_t^{reco} <200 GeV. Different Z masses are shown (500, 750, 1000, 1250 GeV) with the width $\Gamma_{Z'} = 1\% M_{Z'}$.

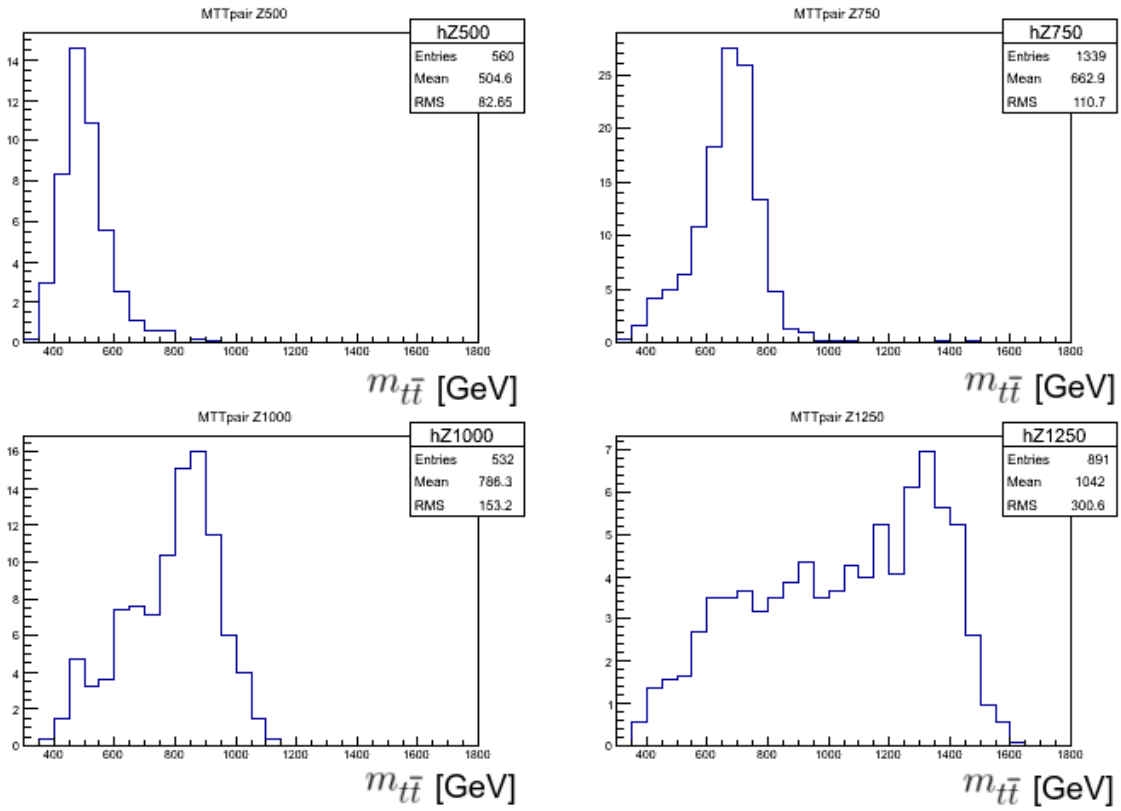


Figure 4.7: $M_{t\bar{t}}^{reco}$ distribution for Z' MC events passing the $\chi^2 < 15$ request and with 150< M_t^{reco} <200 GeV. Different Z masses are shown (500, 750, 1000, 1250 GeV) with the width $\Gamma_{Z'} = 10\% M_{Z'}$.

Mass (GeV)	Width 1% (GeV)	N_{event}
750	7.5	93281
1000	10	103095
1250	12.5	97874
1500	15	97349

Table 4.2: $Z' \to t\bar{t}$ events generated with different masses and widths. Here $\Gamma_{Z'} = 1\% M_{Z'}$.

Mass (GeV)	Width 10% (GeV)	N_{event}
750	75	106587
1000	100	104043
1250	125	85297
1500	150	98775

Table 4.3: $Z' \to t\bar{t}$ events generated with different masses and widths. Here $\Gamma_{Z'} = 10\% M_{Z'}$.

- in the absence of a clear Z′ signal, we derive upper limits on the cross section.

4.2 Systematic uncertainties

The most important sources of systematic uncertainties associated to our selection are:

- simulation of the trigger efficiency;
- knowledge of the jet energy scale (JES) and resolution;
- modelling of the signal (generator);
- choice of the parton distribution functions and QCD scales;
- modelling of the background;
- uncertainty on the integrated luminosity value.
- uncertainty on the b-tagging scale factor.

In the following section we describe three uncertainties that we have considered here: jet energy scale, integrated luminosity and b-tagging.

4.2.1 Jet energy scale

The absolute value of the jet energy scale is unknown but we know instead its associated uncertainty, δ_{JES}, which is expressed as a percentage. To evaluate the uncertainties on the signal yield we calculate first

$$P_\mu^\pm = P_\mu \cdot (1 \pm \delta_{JES}), \tag{4.11}$$

where P_μ is the 4-momentum of the jet and $P_\mu^\pm$ is the modified one, increased/decreased of a factor $(1 \pm \delta_{JES})$. Then the standard selection is applied, obtaining the efficiencies $\epsilon^\pm$ corresponding to $\pm\delta_{JES}$. So, the systematic uncertainty on the jet energy scale is estimated as

$$\frac{|\epsilon_i^+ - \epsilon_i^-|}{\epsilon_i^+ + \epsilon_i^-} = 0.03, \tag{4.12}$$

which is an average uncertainty, applied to every Z′ samples.

4.2.2 Integrated luminosity

The value used for the integrated luminosity has a 2.6% uncertainty, as evaluated by CMS studies.

4.2.3 *b*-tagging

The b-tagging efficiency, as measured in simulations, is different from what expected in data. Such effect is compensated by applying a scale factor. This scale factor has un uncertainty varying between 1.6% and 8% depending on the jet p_T and η. The average value in the range relevant to our selection is about 3% which is doubled to 6% since we require at least two *b*-tags.

4.3 Cross section upper limits

The $M_{t\bar{t}}^{reco}$ spectrum does not show any resonance that can be compatible with the decay $Z' \to t\bar{t}$ predicted in many BSM theories. The Z' production is a rare process and the possibility of not having an observed yield is realistic if the production cross section is small.
Without a observed resonance, the MC simulations for Z' signal allow to compute the upper limit to the Z' production cross section between 500 and 1500 GeV in the $t\bar{t}$ mass spectrum. This result can help to focus future analyses into the search for resonances in the $M_{t\bar{t}}$ distribution.

4.3.1 The package "theta"

The Z' production cross section upper limit is computed using the package "theta"[11], a tool for statistical analysis preferentially dedicated to the HEP; it is usable in C++ and in python and it is compatible with ROOT. For this analysis a python script is written.
Theta allows to fit a model based on the data including the systematic uncertainties on the upper limits in an easier way than with a ROOT code.

4.3.2 Statistical model

The statistical model defined in the theta script is a model with multiple channels (also called "observables") where in each channel c, the model is binned. In every bin i, the number of events follows a Poisson distribution with mean value $\lambda_{c,i}$, that is given by

$$\lambda_{c,i}(\beta_{signal}) = \beta_{signal} S_{c,i} + \sum_{p} B_{c,p,i} \ , \tag{4.13}$$

where

- $S_{c,i}$ is the signal normalized distribution (template) of the i-th bin, scaled to an arbitrary cross section (usually a prediction of the theory, if available);

- $B_{c,i,p}$ is the background prediction template in channel c, of process p, in the bin i, scaled to the cross section and luminosity of the analyzed dataset;

- β_{signal} is what returned from the fit.

If there are systematic uncertainties, they are added as nuisance parameters θ on which the signal and background distributions depend. It is also possible to have many signal templates $S_{s,c,i}$; the Poisson mean in the i-th bin becomes

$$\lambda_{c,i}(\beta_{signal,i}, \theta) = \beta_{signal} \sum_{s} S_{s,c,i}(\theta) + \sum_{p} c_{c,p}(\theta) \cdot B_{c,p,i}(\theta) \tag{4.14}$$

where the backgroung uncertainty is modelled with $c_{c,p}(\theta)$.

Systematic uncertainties and confidence interval

The sistematic uncertainties that affect $\lambda_{c,i}$ are treated modelling θ with a Gaussian distribution with mean value zero and variance equal to the total systematic uncertainty (chapter 4.2). The confidence bands for a nσ confidence level (CL) are obtained shifting by a value nσ the nuisance parameter and finding the corresponding shift for the signal and background templates; this process use a smooth function, which is cubic in the range up to 1σ and a linear extrapolation beyond 1σ, to interpolate between the "nominal" templates (which are not affected by the uncertainty) and the shifted templates.

Theta tecnical execution

Inside theta, the model for $\lambda_{c,i}$ is created with the function **build_model_from_rootfile** declared in the python class **Model**. The Model class contains all relevant information about the model, including the observed data, and all the predictions including their dependence on the model parameters in the different channels, and which of the processes is to be considered as signal.
The systematic uncertainties are added to a defined process calling from Model the function **add_lognormal_uncertainty** and inserting the relative uncertainty.
The evaluation of the model for every signal process is executed with randomly generated distributions called "toys".

$\lambda_{c,i}$ for $\sigma_{Z'}$ upper limit

Our analysis uses a function

$$\lambda_{c,i}(\beta_{signal,i}, \theta) = \beta_{signal} S_{c,i}(\theta) + \sum_{p} c_{c,p}(\theta) \cdot B_{c,p,i}(\theta), \qquad (4.15)$$

where the processes p are the $t\bar{t}$ and the multijet background simulations, and in this case are both treated as background processes for Z' production.

4.3.3 The Theta analysis

Using an auxiliar ROOT macro, the theta input files were created. These files contain all the histograms useful for the upper limit computing with a fixed χ^2 cut and the Z' predictions with a certain choise for $M_{Z'}$ and $\Gamma_{Z'}$.

Then the python script builds a model based on the $t\bar{t}$ simulation and the multijet background including the systematic uncertainties, that are

- $\sigma_L = 2.6\%$ on the *integrated luminosity*;
- $\sigma_{JES} = 3\%$ on the *jet energy scale* (JES);
- $\sigma_{b-tag} = 6\%$ on the *b-tagging efficiency.*

The cross section upper limit is computed using a Bayesian likelihood-based method, allowing the expected background model ($t\bar{t}$ and multijet events) and the five Z′ signals to fluctuate within the various systematic and statistical uncertainties to find the best fit to the observed data distribution. The results are shown in figures 4.8 and 4.9 and summarized in Tables 4.4 ad 4.5. The black solid line represents the median expected limit at 95%, while the dark green and light green bands represent the 1σ and 2σ bands, respectively, of the expected limits from the set of pseudoexperiments. The red solid line shows the observed limit results using the observed data distribution. The blue dashed line shows the theoretical cross sections curves that comes from the NLO calculation in the Leptophobic topcolor Z' model [10]. Comparing the theoretical curves and the observed limits, we can set exclusion limits on the $M_{Z'}$ mass for each choise of $\Gamma_{Z'}$. The limits are set where the two curves intersect each other and are:

$$M_{Z'} > 1300 \text{ GeV } (\text{for } \Gamma_{Z'} = 1\% M_{Z'})$$

$$M_{Z'} > 1500 \text{ GeV } (\text{for } \Gamma_{Z'} = 10\% M_{Z'}) \ .$$

$M_{Z'}$[GeV]	$\Gamma_{Z'}(1\%)$[GeV]	$\sigma^{meas}_{Z'}$[pb]	$\sigma^{exp}_{Z'}$[pb]				
			-2σ	-1σ	median	1σ	2σ
500	5	5.92	1.89	2.52	3.53	4.97	6.75
750	7.5	0.43	0.46	0.62	0.86	1.22	1.65
1000	10	0.24	0.31	0.41	0.58	0.82	1.12
1250	12.5	0.34	0.32	0.43	0.61	0.87	1.2
1500	15	0.55	0.37	0.51	0.73	1.06	1.49

Table 4.4: Expected and observed upper limits on the Z′ production cross section, with intervals corrisponding to 68% and 95% CL for different masses and width $\Gamma_{Z'} = 1\% M_{Z'}$.

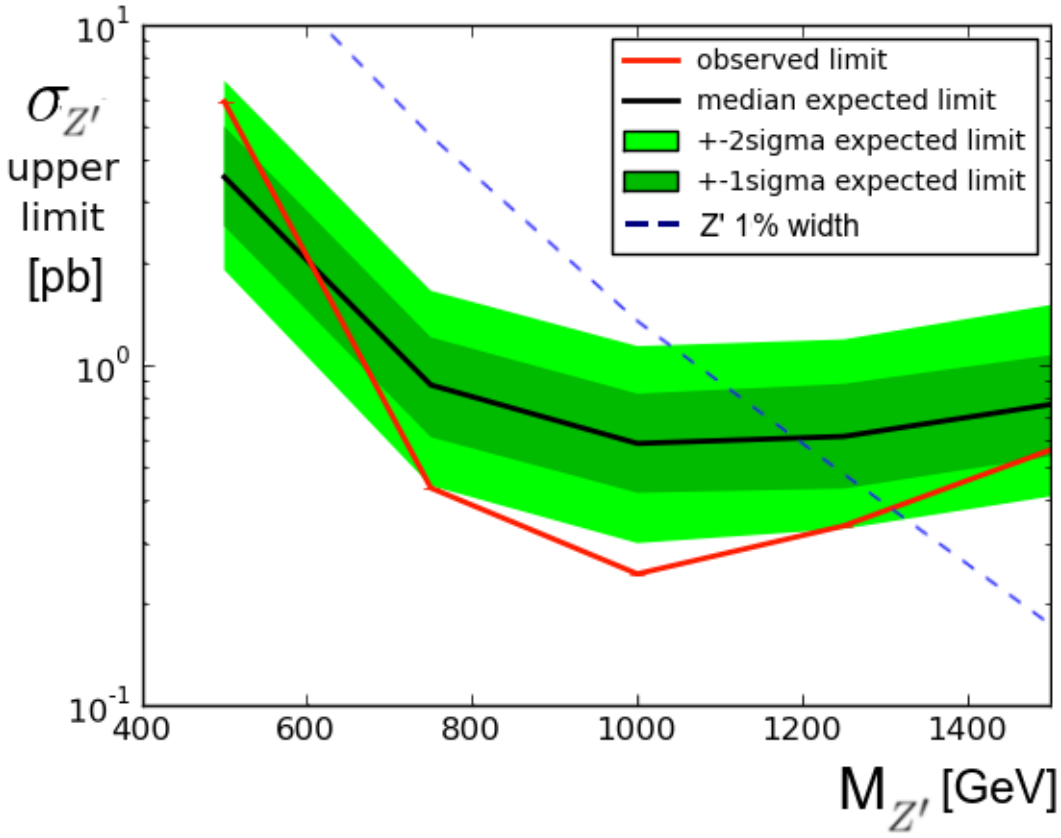


Figure 4.8: Upper limit on the Z' production cross section for $\Gamma_{Z'}/M_{Z'} = 1\%$.

$M_{Z'}$[GeV]	$\Gamma_{Z'}(10\%)$[GeV]	$\sigma^{meas}_{Z'}$[pb]	$\sigma^{exp}_{Z'}$[pb]				
			-2σ	-1σ	median	1σ	2σ
500	50	5.12	1.95	2.6	3.63	5.11	6.94
750	75	0.26	0.34	0.45	0.63	0.90	1.22
1000	100	0.33	0.42	0.56	0.79	1.12	1.53
1250	125	0.50	0.49	0.65	0.92	1.31	1.81
1500	150	0.26	0.20	0.27	0.39	0.56	0.79

Table 4.5: Expected and observed upper limits on the Z′ production cross section, with intervals corrisponding to 68% and 95% CL for different masses and width $\Gamma_{Z'} = 10\% M_{Z'}$.

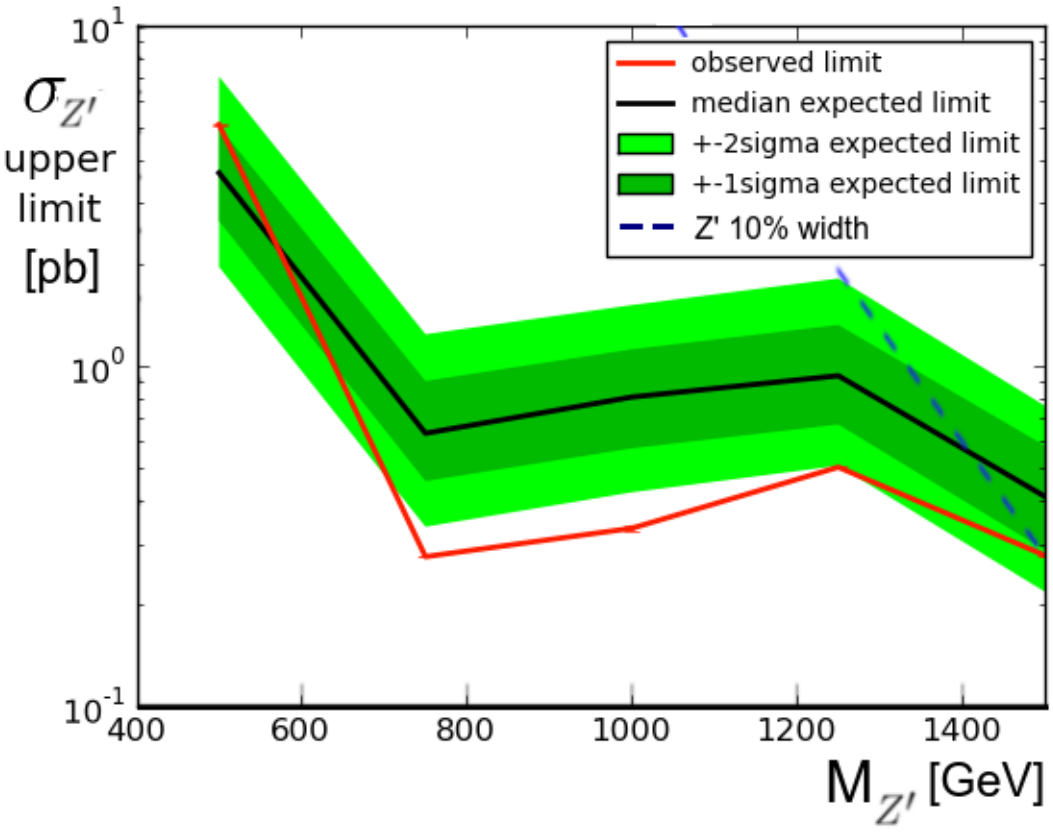


Figure 4.9: Upper limit on the Z' production cross section for $\Gamma_{Z'}/M_{Z'} = 10\%$.

Conclusions

The search for resonances on the $M_{t\bar{t}}$ spectrum is crucial for the study of the physics beyond SM.
First of all, we applied an event selection to discriminate $t\bar{t}$ events from background. We used cuts on the p_T and the $|\eta|$ of every jet, on the number of b-tagged jets of each event and their ΔR_{bb} separation. The signal we searched for is the $t\bar{t}$ all-hadronic final state of a Z$'$ resonance, that is

$$pp \to Z' \to t\bar{t} \to q_1\overline{q_2}q_3\overline{q_4}b\bar{b} \ .$$

This event selection favoured the purity of the events that are passing the cuts, because the Z$'$ events are very rare. The requests were: $p_T > 30$ GeV, $|\eta| < 2.4$, $N_{b-tag} \geq 2$ and $\Delta R_{bb} > 1.5$.
After the event selection, the kinematic reconstruction based on a χ^2 function returned with a likelihood fit, the $M_{t\bar{t}}$ MC distributions for $t\bar{t}$, background and the data.
Then, the same approach is used on the Z$'$ MC simulations for different $M_{Z'}$ (500, 750, 1000, 1250 and 1500 GeV) and width Γ_Z (1% and 10% of the Z$'$ mass).
We observed that there was no evidence for resonances at masses between 500 and 1500 GeV in the $t\bar{t}$ mass distribution; instead, an agreement with the expectation from the SM is seen.
Finally, we calculated the cross section upper limits for Z' production using the statistical tool "theta". This allowed us to put lower limits on the Z$'$ mass: $M_{Z'}(\Gamma_{Z'} = 1\%M_{Z'}) > 1300$ GeV and $M_{Z'}(\Gamma_{Z'} = 10\%M_{Z'}) > 1500$ GeV.

Bibliography

[1] Alioli, Simone and others, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP 06 (2010) 043, doi:10.1007/JHEP06(2010)043.

[2] J.Beringer et al.*Review of Particle Physics*, Phys. Rev. D **86**, 010001, 2012.

[3] CMS Collaboration - *Particle-flow event reconstruction in CMS and performance for jets, taus and emiss*, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[4] CMS Collaboration - *Commissioning of the Particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector*, CMS Physics Analysis Summary CMS-PAS-PFT-10-001, 2010.

[5] M.Cacciari, G. P. Salam and G. Soyez - *The anti-k_T jet clustering algorithm*, JHEP 04 (2008) 063.

[6] M.Cacciari, G. P. Salam and G. Soyez - *FastJet user manual*, Eur. Phys. J.C 72 (2012) 1896.

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

[8] M.Kaku, *Quantum Field Theory: A Modern Introduction* (New York, USA:Oxford University Press, 1993).

[9] M. Cvetic & S. Godfrey *Discovery and identification of extra gauge bosons* (1995). Hep-ph/9504216.

[10] Robert M. Harris, Supriya Jain - *Cross Sections for Leptophobic Topcolor Z' decaying to top-antitop* (arXiv: 1112.4928v3)

[11] T. Muller , J. Ott, J. Wagner-Kuhr - *theta - a framework for template-based modeling and inference*, 2010, http://www-ekp.physik.uni-karlsruhe.de/ ott/theta/theta-auto/

List of Figures

1.1 The fundamental particles of the Standard Model. 8
1.2 Representation of the pp collision in the parton model. 9
1.3 Evolution of the SM couplings $\alpha_i = \frac{g_i^2}{4\pi}$ as a function of the energy scale. 11
1.4 Branching Ratios for the decay channels of the $t\bar{t}$ pair. 15

2.1 The LHC and the positions of the 4 main detectors: ATLAS, CMS, ALICE and LHCb. 18
2.2 Scheme of the Large Hadron Collider and of the other acceleration rings. 19
2.3 The CMS detector. 19
2.4 Section view of the CMS experiment. Different particles show different behaviours and trajectories within the various subdetectors. 21

3.1 Feynman diagram for $gg \to Z' \to t\bar{t}$ process, with the all-hadronic final state. 25
3.2 CVS b-tagging discriminator for $t\bar{t}$ MC events (a) and data, i.e. generic multijet events (b) 30
3.3 ΔR_{bb} distribution for the $t\bar{t}$ MC events (blue line) and the data (red line). 31

4.1 χ^2 distributions for multijet data (a) and $t\bar{t}$ MC events (b). 37
4.2 M_t^{reco} distributions for $t\bar{t}$ events (a) and the multijet background events(b), with $\chi^2 < 15$. 38
4.3 $M_{t\bar{t}}^{reco}$ distributions for $t\bar{t}$ events (a) and the multijet background events (b), with $\chi^2 < 15$. 39
4.4 M_t^{reco} distributions for $\chi^2 < 15$ (a) and $\chi^2 < 5$ (b). The $t\bar{t}$ and background yields are normalized to the fit outcome. 42

4.5 $M_{t\bar{t}}^{reco}$ for $t\bar{t}$ events and for multijet events with $\chi^2 < 15$ and 150¡M_t^{reco}¡200 GeV. Shown are MC distributions for several Z' simulations with different masses and with different width: $\Gamma_{Z'} = 10\%$ (a) and $\Gamma_{Z'} = 1\%$ (b), assuming a production cross section of 1 pb. 43
4.6 $M_{t\bar{t}}^{reco}$ distribution for Z' MC events passing the $\chi^2 < 15$ request and with 150< M_t^{reco} <200 GeV. Different Z masses are shown (500, 750, 1000, 1250 GeV) with the width $\Gamma_{Z'} = 1\% M_{Z'}$. 44
4.7 $M_{t\bar{t}}^{reco}$ distribution for Z' MC events passing the $\chi^2 < 15$ request and with 150< M_t^{reco} <200 GeV. Different Z masses are shown (500, 750, 1000, 1250 GeV) with the width $\Gamma_{Z'} = 10\% M_{Z'}$. 45
4.8 Upper limit on the Z' production cross section for $\Gamma_{Z'}/M_{Z'} = 1\%$. 51
4.9 Upper limit on the Z' production cross section for $\Gamma_{Z'}/M_{Z'} = 10\%$. 52

List of Tables

1.1 Quantum numbers for the SM fermions 6
1.2 Theoretical predictions for $\sigma_{Z'} B(Z' \to t\bar{t})$ calculated with the Leptophobic Topcolor Model at LO for different masses and width. . . 16

4.1 Fractions and number of $t\bar{t}$ and background events from χ^2 fit. . . 41
4.2 $Z' \to t\bar{t}$ events generated with different masses and widths. Here $\Gamma_{Z'} = 1\% M_{Z'}$. 46
4.3 $Z' \to t\bar{t}$ events generated with different masses and widths. Here $\Gamma_{Z'} = 10\% M_{Z'}$. 46
4.4 Expected and observed upper limits on the Z′ production cross section, with intervals corrisponding to 68% and 95% CL for different masses and width $\Gamma_{Z'} = 1\% M_{Z'}$. 50
4.5 Expected and observed upper limits on the Z′ production cross section, with intervals corrisponding to 68% and 95% CL for different masses and width $\Gamma_{Z'} = 10\% M_{Z'}$. 51