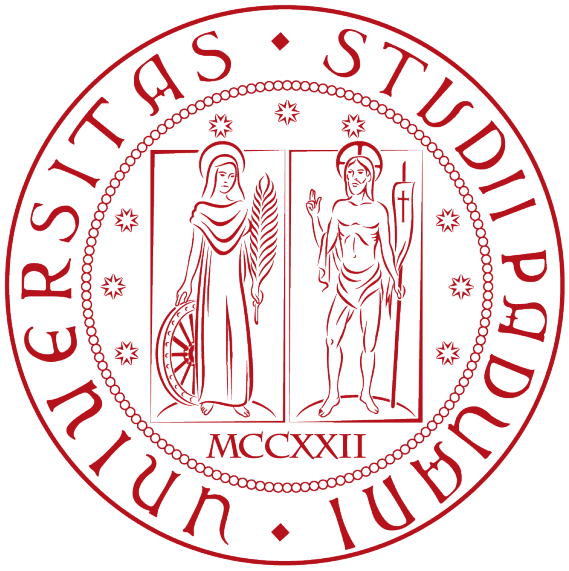
Università degli Studi di Padova

Dipartimento di Fisica e Astronomia
"Galileo Galilei"

Corso di Laurea in Fisica



Search for associated production of Higgs boson and top-quark pairs with CMS

Laureando
Matteo Migliorini

Relatore
Dott. Tommaso Dorigo

Anno Accademico 2016-2017

Do. Or do not. There is no try.
— Yoda

Contents

Abstract

At the LHC, the decay signal of top quark pairs produced in association with a Higgs boson ($t\bar{t}H$), when all final state bodies decay into jet pairs, is deeply concealed within a huge background of inclusive top quark pair production. In this work we try to distinguish the two processes by directly associating each final state parton to an observed hadronic jet, through the minimisation of a χ^2 statistic. Three possible choices of the statistic are investigated using simulated samples of the relevant process. Results of the various choices are shown and compared.

Chapter 1

Introduction

1.1 The CMS experiment at the LHC

1.1.1 The LHC

The *Large Hadron Collider* (LHC) is currently the world's largest and most powerful particle accelerator. It first started up on 10 September 2008, and remains the latest addition to the CERN's accelerator complex (Figure 1.1). Located between

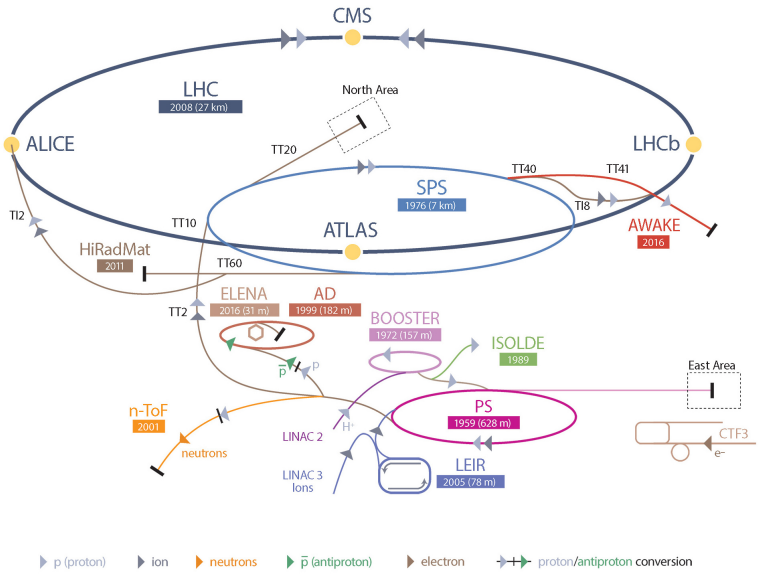


Figure 1.1: Map of the CERN accelerator complex

France and Switzerland, it consists of a 27 km ring placed in an underground tunnel were two opposite beams of protons (or heavy ions), travelling into two different

beam pipes, are forced to collide at the center of the four experiments (ALICE, ATLAS, CMS and LHCb). In order to keep the particles inside the ring, the LHC is endowed whit 1232 superconducting dipole magnets that bend the beams, and 392 superconducting quadrupole magnets which focus the beams. Another tipe of magnet is used to squeeze the particle beams before the collision, in order to increase the chances of collision. The magnets, in order to operate in a superconducting state, are kept at a temperature of 1.9 K using liquid helium.
LHC is designed to accelerate beams of protons up to a center of mass energy of $\sqrt{s}$ =14 TeV and a luminosity of $1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The luminosity goal was reached in June 2016 and the design energy will be reached after a future upgrade of the accelerator.

In order to reach the design energy, protons are accelerated by a chain of accelerators (shown in Figure 1.1). Protons (obtained by the ionisation of hydrogen atoms) are injected in LINAC 2, which accelerate them up to an energy of 50 MeV. The beam coming from LINAC 2 is then accelerated to 1.4 GeV with the *Proton Synchrotron Booster* (PSB). The beam from PSB is injected into the *Proton Sincrotron* (PS) were protons reach an energy of 28 GeV (designed to separate the bunches by 25 ns) and then they reach the energy of 450 GeV in the *Super Proton Synchrotron* (SPS). Lastly, protons are injected in the LHC, where they reach the final energy of 6.5 TeV.

Once the final energy has been reached, the two beams collide in four points of the ring in correspondence of the four experiment: *Compact Muon Solenoid* (CMS), *A Toroidal LHC ApparatuS* (ATLAS), *A Large Ion Collider Experiment* (ALICE) and *Large Hadron Collider beauty* (LHCb). ATLAS and CMS are general purpose detectors bult in order to test the standard model and discover the Higgs boson, ALICE is dedicated to the study of the heavy ions collision and LHCb is dedicated to the study of the bottom quark.

1.1.2 The CMS experiment

The *Compact Muon Solenoid* (CMS) is one of the two general-purpose detectors at LHC (the other one is ATLAS) and is located in one of the four LHC collision points. It has a broad physics programme ranging from studying the Standard Model (remarkable is the discover of the Higgs boson, announced on 4 July 2012 by the ATLAS and CMS collaborations [1] [2]) to dark matter and Supersymmetry. The CMS detector has a cylindrical shape symmetrical around the beam pipe, with a diameter of 14.6 m, a total length of 22 m and a weight of 12500 tons. It is divided into a central section, made of several layers coaxial to the beam axis (the *Barrel*), closed at its end by two hermetic discs orthogonal to the beam (the *Endcaps*). A schematic view of a transverse section of CMS is shown in Figure 1.2. Starting from the beam position, and moving outwards, it presents a silicon tracker, a electro-

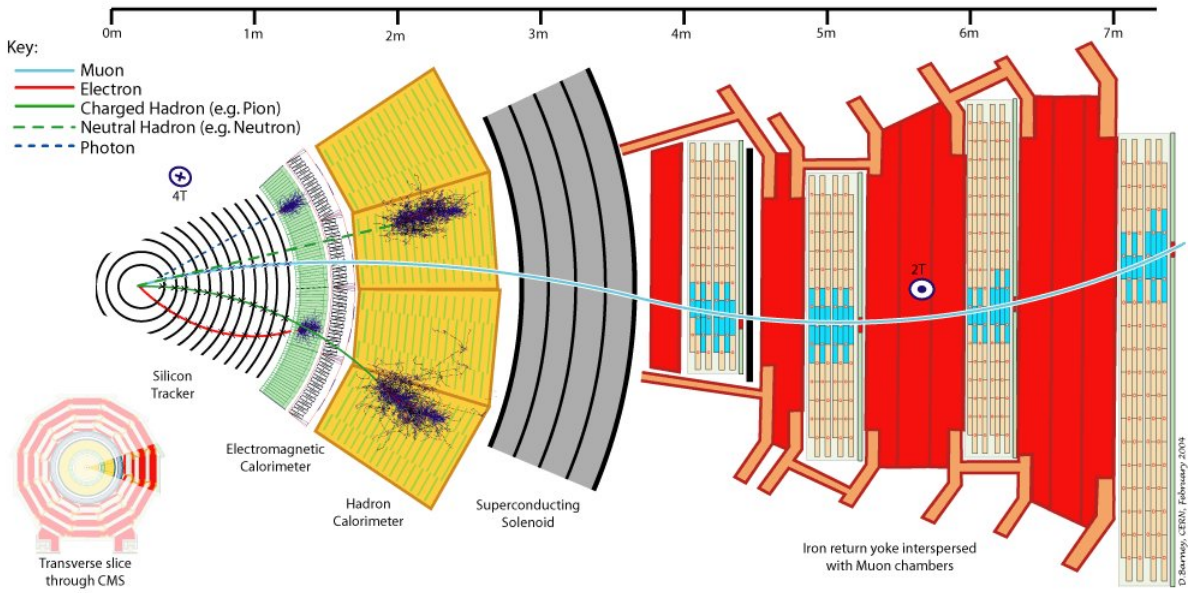


Figure 1.2: Schematic view of the transverse slice of CMS

magnetic calorimeter(ECAL), a hadronic calorimeter (HCAL), the superconducting solenoidal magnet and the muon drift chamber in the return yoke.
The coordinate system adopted in CMS is cartesian, it has the origin centred in the collision point at the center of the detector and adopts the following conventions:

- The x axis points toward the center of the LHC center;
- The y axis points towards the surface;
- The z axis is parallel to the beam pipe.

The cylindrical symmetry of the detector suggests to use polar coordinates (r, ϕ, θ) where r is the radial distance from the collision point, ϕ is the azimuthal angle with respect to the x axis and θ is the polar angle with respect to the z axis. Instead of using θ CMS use the *pseudorapidity* defined as

$$\eta = -\ln \tan \frac{\theta}{2} \tag{1.1.1}$$

In hadron colliders rapidity is preferred over the polar ange θ because particle production is constant as a function of rapidity and because differences in pseudorapidity $\Delta\eta$ are Lorentz invariant under boosts along the longitudinal axis, thus simplifying the reconstruction of hadronic jets which retain a conical shape for any pseudorapidity.
We now give a brief description of the magnet and of the main CMS subdetectors[3][4].

Magnet The large superconducting magnet of CMS is a solenoid that delivers an axial and uniform magnetic field of 3.8 T over a length of 12.5 m and a radius of 3.5 m. Inside the magnet there are the tracker, the ECAL and the HCAL. This is advantageous because it allows to minimise the energy losses before the calorimeter due to the interaction of particles with the coil.

Silicon tracker The silicon tracker is a cylinder-shape detector with a length of 5.6 m and an outer radius of 1.20 m. The barrel is formed of 3 layers of silicon pixel detectors, surrounded by 10 layers of silicon micro-strip detectors while the two endcaps comprise 2 layers of silicon pixel detectors and 12 layers of silicon micro-strip detectors. The spatial resolution is about 10 µm and it allows to reconstruct precisely the interaction vertex and secondary vertex.

Electromagnetic calorimeter The ECAL is a hermetic and homogenous calorimeter made of lead tungstate ($PbWO_4$) crystals. The barrel covers $|\eta| < 1.479$ and the two endcaps disks $1.479 < |\eta| < 3.0$. Crystals are long enough (in term of radiation length X_0) to contain 98% of the energy of photons and electrons with a energy up to 1 TeV ($L^{\text{barrel}} = 25.8\,X_0$ and $L^{\text{endcaps}} = 24.7\,X_0$). The crystals transverse size matches the Molière radius R_M =2.2 cm. The fine transverse granularity makes it possible to fully resolve hadron and photon energy when the particles are as close as 5 cm from another, necessary in order to reconstruct the correct jet energy. Crystals has a very fast light emission (5 ns for the principal scintillation and 15 ns for the secondary scintillation) and this light is collected by avalanche photodiodes (APD) in the barrel and vacuum phototriodes (VPT) in the barrel.

Hadronic calorimeter The HCAL is a hermetic sampling calorimeter, formed of several layers of brass absorber and plastic scintillator. It surrounds ECAL and covers $|\eta| < 1.3$ with the barrel and $1.3 < |\eta| < 3.0$ with two endcaps disks. The thickness (in terms of interaction length λ_I) goes from $7\lambda_I$ to $11\lambda_I$, depending on η. The scintillation light is collected an converted by wavelength-shifter fibres (WLS) and this light is read by hybrid photodiodes (HPD). The HCAL is completed by the hadronic forward (HF) calorimeters, situated at ± 11 m from the interaction point that extend the the coverage up to $|\eta| \sim 5$. It consists of steel absorber composed of groove plates with quartz fibres inserted in the grooves and read by photomultipliers.

Muon detector Outside the solenoid coil, the magnetic returned trough a yoke consisting of three layers of steel interleaved with four muon detector planes. Drift tubes (DT) detect muons in the region $|\eta| < 1.2$ and cathode strip chambers (CSC) in the region $0.4 < |\eta| < 2.9$. They are complemented by resistive plate chambers (RPC) covering the range $\eta < 1.6$.

1.1.3 Jet Reconstruction

A jet is a narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon due to the fact that coloured particles only exist in a colour-neutral state, such as meson and baryons. The aim of jet reconstruction is to measure the momentum of the coloured parton which initiated the hadronization. Jets are composed by many particles, therefore we need a clustering algorithm capable to univocally define it. Those algorithms must satisfy two requirements:

- *collinear safety:* if a particle of momentum p is substituted by two collinear particles of momentum $p/2$ the result of the algorithm must be the same;
- *infrared safety:* if an infinitely soft gluon is added to the list of particles which have to be clustered, the result of the algorithm must be the same.

The *anti-k_T*[5] algorithm is the jet algorithm employed at CMS. It proceeds in this way:

- define the distance d_{ij} between two particles i and j as

$$d_{ij} = \min\left(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}\right)\frac{\Delta R_{ij}^2}{R^2}$$

  where $p_{T_{i,j}}$ are the particles transverse momenta, $\Delta R_{ij} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is the distance in the (η, ϕ) plane. Define also the distance between a particle i and the beam as

$$d_{iB} = \frac{1}{p_{T_i}^2}$$

- find the minimum of all d_{ij} and d_{iB}
  - if smallest is a d_{iB}, remove the particle i and call it a jet;
  - if smallest is a d_{ij}, combine (sum of four-momenta) the two particles i and j, update the distances and then proceed finding the next smallest.
- iterate until only jets are left.

We can easily see that a jet cannot contain particles at a distance greater than R from the central axis, therefore jets are cone-shaped. The jets used in this thesis have been found using *anti-k_T* with a radius parameter R=0.5.

At reconstruction level, a list of particle candidates as jet is passed to the algorithm, produced by the event reconstruction technique known as *Particle Flow* (PF) [4]. The Particle Flow algorithm identifies and reconstructs individual particles in the event by combining the information from all the subdetectors.

1.1.4 b-tagging

The identification of b-jets (*b-tagging*) is important to study top quarks, as well as Higgs boson in the dominant $b\bar{b}$ decay mode (in this thesis $t \to W^+b$ and $H \to b\bar{b}$ were considered). The jets originated from a b quark can be identified thanks to

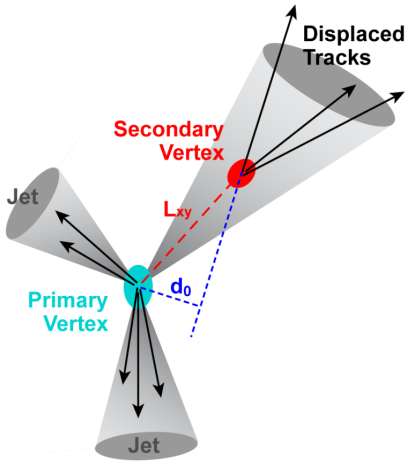


Figure 1.3: Representation of a B-hadron decay in the (x,y) plane

the long lifetime of the b quark $c\tau$ ~450 µm, so a B-hadron with p_T=50 GeV covers almost half a centimetre before decaying (L_{xy} in Figure 1.3). Thus the particle coming from the b-quark decay can have a sizeable impact parameter whit respect to the primary vertex (d_0 in Figure 1.3). The b-tagging algorithm is based on the measure of three main variables: the impact parameter, the position of the secondary vertex and the transverse momentum of the muon relative to the jet direction[6].

A variety of algorithms have been developed by the CMS collaboration to identify b-jets based on the B-hadron properties just described. In this thesis CSV algorithm has been considered. The *Combine Secondary Vertex* algorithm (CSV) is a vertex-based algorithm which collects in a single discriminator the information provided by all the possible variables. A likelihood discriminator is built and trained on the three categories using the following set of variables:

- the vertex category ("real", "pseudo" or "no vertex");
- the flight distance significance in the transverse plane ("2D");
- the vertex mass;
- the number of tracks in the vertex;

- the ratio of energy carried by tracks at the vertex with respect to all tracks in the jet;
- the 2D IP significances (Impact parameter d_0 significance=$d_{0,Sign.} = d_0/\sigma_{d_0}$) of the first track that raises the invariant mass above the charm threshold $1.5\,\mathrm{GeV}/c^2$;
- the number of tracks in the jet;
- the 3D IP significances for each track in the jet.

Two likelihood ratios are built from these variables. They are used to discriminate between b and c jets and between b and light partons. They are then combined with prior weights of 0.25 and 0.75 respectively. The CSV algorithm classifies the tag according to their discriminator value as *Loose* (≥ 0.5426), *Medium* (≥ 0.8484) or *Tight* (≥ 0.9535). The tighter the selection is, the more likely the jets are truly coming from a b quark, but the efficiency decreases.

1.2 Standard Model and Higgs boson

1.2.1 The Standard Model

The Standard Model (SM) of particle physics is the theory describing three of the four known fundamental forces (*Strong*, *electromagnetic* and *weak* interactions), as well as classifying all known elementary particles. It was developed throughout the second half of the 20th century and it has reached his current formulation in the mid-1970s thanks to the work of *S. Glashow*, *S. Weinberg* and *A. Salam*.
As it is possible to see in Figure 1.4 the Standard Model divides all the elementary particles into two main categories: *fermions*, particles with half-integer spin, and *bosons*, particle with integer spin. In the SM there are two types of elementary fermions: quarks (*up*, *down*, *charm*, *strange*, *top* and *bottom*) and leptons (*electron*, *muon*, *tau*, *electron neutrino*, *muon neutrino* and *tau neutrino*). The elementary bosons are all gauge bosons, the carriers of the forces (except the Higgs boson which is a scalar boson): the *photon* (carrier of the electromagnetic force), the *gluon* (carrier of the strong force), the Z and W bosons (carrier of the weak force).

1.2.2 The Higgs boson

The Higgs boson is a massive scalar elementary particle theorised by Peter Higgs in 1964. It is a key building block in the Standard Model and has no intrinsic spin (for that reason is classified as a boson). The Higgs Boson explains why the other elementary particles, except the photon and the gluon, are massive. Due to its high

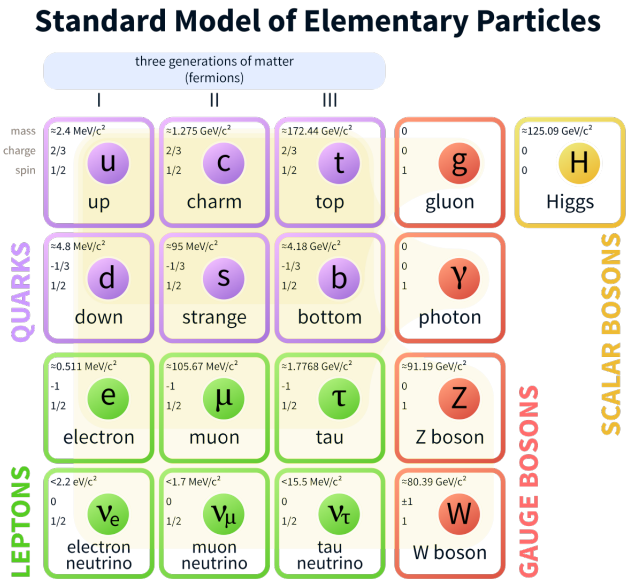


Figure 1.4: The Standard Model of elementary particles

mass and the fact that it decay almost immediately when created, it requires a high energy accelerator to be observed. On 4 July 2012, ATLAS and CMS reported independently the observation of a new particle with a mass of about 125 GeV[1][2] consistent with the SM Higgs boson.

Role of the Higgs boson in the standard model

Since the electroweak interaction breaks the chiral symmetry and the Lagrangian of a particle must satisfy the gauge invariance of the group SU(3)xSU(2)xU(1), the term in the Lagrangian that contains the mass of the particle vanishes, therefore the particle remains massless. The solution of this problem is to consider a complex scalar field H (with a potential shaped as it is possible to see in Fig. 1.5) which maintains a symmetrical structure but the stable states are in the lowest potential region where the field reaches its "expectation vacuum value" and are symmetrical. Because the value of the potential at H=0 must be 0, the vacuum expectation value is $v \neq 0$ thus the field equation can be written as $h(\phi) = v + H(\phi)$. It can be shown that the Lagrangian density is

$$\mathcal{L} = \partial_\mu h^* \partial^\mu h - m^2 h^* h - \lambda (h^* h)^2 - \lambda_f h f^* f - g^2 W^+ W^- h^2 - \rho^2 \left(Z^0\right)^2 h^2 \quad (1.2.1)$$

where λ is the coupling constant, f the wave function of a fermion, m the mass of the Higgs boson and λ_f, g^2, ρ^2 are respectively the coupling constants of Higgs-fermion coupling, Higgs-W and Higgs-Z couplings. The term $\lambda_f v f^* f$ can be schematised as a

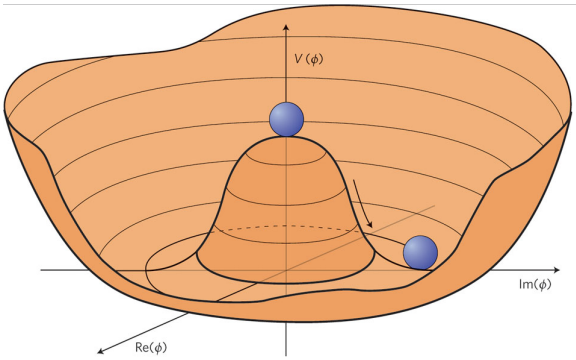


Figure 1.5: Higgs field potential V as a function of the phase ϕ of the field.

Feynman diagram with 2 vertices and $\lambda_f v$ as the coupling constant, which represent the mass of the fermion. This happens also for the W and Z boson. This generation of the mass is called *Higgs mechanism*.

The Higgs boson production and decay channels

The Higgs boson production cross section at the LHC depends on the energy of the beams and the Higgs boson mass value (we assume here m_H=125 GeV, which is very close to the current measured value by ATLAS and CMS [7]). In proton-proton collisions, there are four mechanisms of production involved: gluon-gluon fusion (ggH), vector-boson fusion (VBF), associate production with vector boson (VH) and associated production with top or bottom quarks (ttH/bbH). Higgs production cross sections are shown in the Table 1.1[8].

	σ [pB]	σ/σ_{tot}
ggH	43.92	86.2%
VBF	3.748	7.4%
WH	1.380	2.7%
ZH	0.8696	1.7%
$b\bar{b}$H	0.5116	1.0%
$t\bar{t}$H	0.5085	0.9%

Table 1.1: Higgs boson production cross section at $\sqrt{s}$=13 TeV assuming m_H=125 GeV.

In hadron collisions the Higgs boson can be detected only through its decay products. The various decay channels present a different decay rate depending on

the coupling to the Higgs boson. The branching ratios (BR) are defined as:

$$BR\left(H \to a\right) \equiv \frac{\Gamma\left(H \to a\right)}{\Gamma_H} \tag{1.2.2}$$

where $\Gamma\left(H \to a\right)$ is the inverse decay rate of the final state a and Γ_H the sum over all the allowed decay channels. The Higgs boson branching ratios are reported in Table 1.2.

	$b\bar{b}$	WW	gg	$\tau\tau$	$c\bar{c}$	ZZ	$\gamma\gamma$	$Z\gamma$	$\mu\mu$
BR(%)	57.5	21.6	8.56	6.3	2.9	2.7	0.23	0.15	0.02

Table 1.2: Higgs boson decay Branching Ratios assuming m_H=125 GeV.

1.2.3 $t\bar{t}H$ production

In this thesis the associated production of top-antitop quark pair and a Higgs boson has been considered (Figure 1.6).

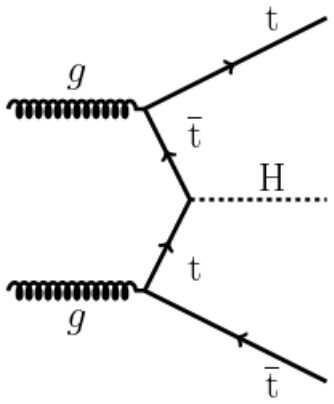


Figure 1.6: Feynman diagram for $t\bar{t}H$ production.

This channel, as it is possible to see in Table 1.1, is the one with the lowest cross section (at $\sqrt{s}$=13 TeV) among all lowest-order production processes of Higgs bosons at the LHC, a factor 100 smaller than the dominant gluon-gluon fusion (ggH) production in pp→H production. The observation is interesting because the rate of this process depends on the coupling of the Higgs boson to the top quark. Because of the large mass of the top quark (∼173 GeV) the relative Yukawa coupling is $\lambda_t \sim 1$. The $t\bar{t}H$ production represent the only opportunity to measure directly the top-Higgs coupling without making assumptions about possible contributions from sources beyond the standard model, indeed indirect loop measurement (Figure 1.7)

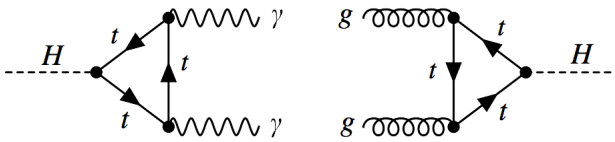


Figure 1.7: Possible way to indirectly measure the top-Higgs coupling.

can be influenced by BSM physics.
Measuring a $t\bar{t}H$ ratio $\mu=\sigma_{ttH}/\sigma_{SM}$ (where σ_{ttH} is the measured cross section and σ_{SM} the one expected from the Standard Model) different from one could indicate the presence of new physics.

$t\bar{t}H(H\rightarrow b\bar{b})$

As it is possible to see in Table 1.2 the $H\rightarrow b\bar{b}$ decay has the largest Higgs branching ratio but the resulting experimental signature is drowned in an irreducible $t\bar{t}+b\bar{b}$ background. The top quarks decay decays with BR>0.99 as $t\rightarrow Wb$; then the W bosons can decay $W\rightarrow q\bar{q}$ or $W\rightarrow l\nu$ with BR respectively equal to 67.60±0.27% and 10.80±0.09% (Figure 1.8).

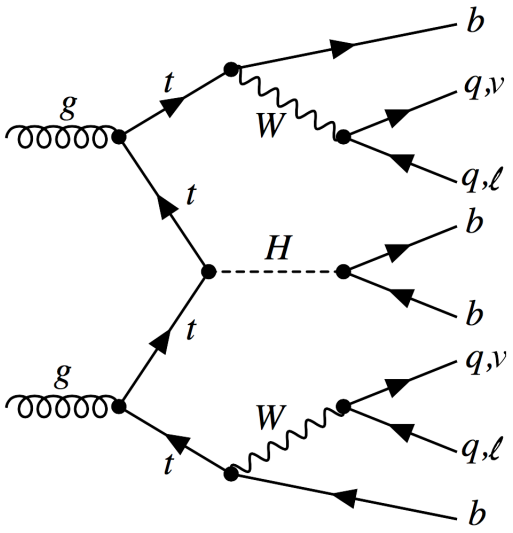


Figure 1.8: $t\bar{t}H$ in the hadronic and leptonic final state.

It is then possible to consider a leptonic, semi-leptonic or fully hadronic final state for the results of the top pair decay. The advantage of the leptonic channel is that one (or two) leptonic decay of the tops (muons or electrons)[1] kills the QCD multijet

[1]The τ are not considered due to their less direct experimental segnature.

background. On the other hand all hadronic $t\bar{t}H$ decays provide access to a large number of events, indeed BR(all hadronic)=46%. The fully hadronic final state has 8 jets: 4 b-jets and 4 light jets (Figure 1.9) and this leads to a large QCD background.

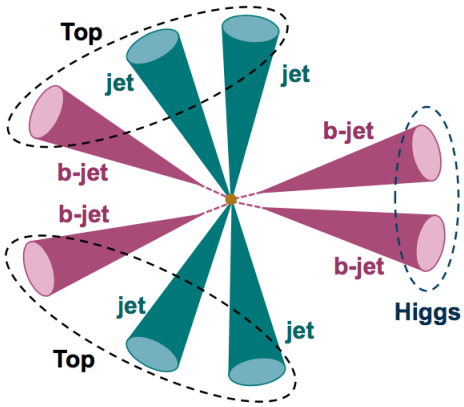


Figure 1.9: $t\bar{t}H$ decay in fully hadronic final state.

The latest results in $t\bar{t}H(H\rightarrow b\bar{b})$ achieved by CMS, measuring the signal strength relative to the standard model cross section $\mu=\sigma/\sigma_{SM}$, observe (expected) upper limit of $\mu<1.5(1.7)$ at 95% C.L. and obtain a best fit value of $\mu = -0.19^{+0.45}_{-0.44}(\text{stat.})^{+0.66}_{-0.68}(\text{syst.})$ (Figure 1.10) [9]. These results are compatible with SM expectations at the level of 1.5 standard deviations.

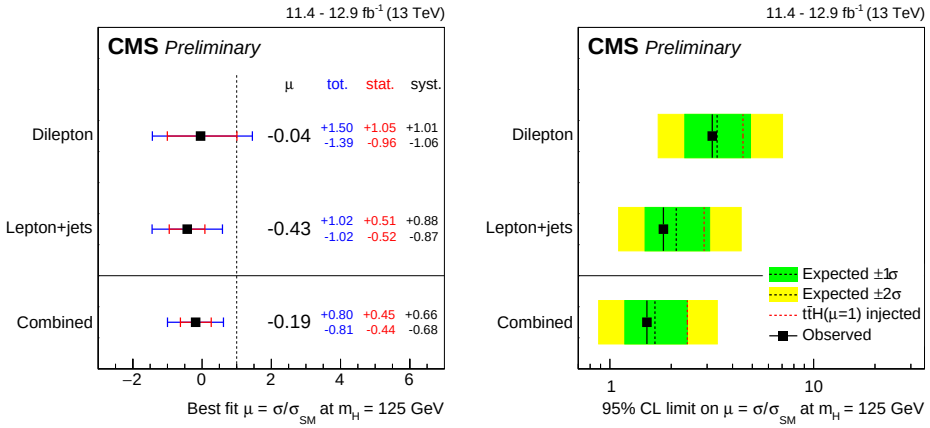


Figure 1.10: Best-fit values of the signal strength modifiers μ (left) and median expected and observed 95% CL upper limits on μ (right)[9].

Chapter 2

Setup of the analysis and preliminary selection

2.1 Data

The simulated samples used in this thesis are composed by events of $t\bar{t}H$ with $H \to b\bar{b}$ and inclusive decay of $t\bar{t}$ ($t\bar{t} \to bW^+\bar{b}W^-$, which includes the hadronic $W^+W^- \to 4q$, leptonic $W^+W^- \to l\bar{\nu}_l\bar{l}\nu_l$ and semi-leptonic $W^+W^- \to q\bar{q}\bar{l}\nu_l$ decay of the W boson). The production of $t\bar{t}H$ is simulated using *Powheg-Box*[10] as generator. The hadronization is handled by *Pythia 8*. The response of the CMS detector is modelled using *Geant 4*[11] and then jets are reconstructed using the *Anti-k_T* algorithm as described in section **1.1.3**. These data are collected in a ROOT *nTuple* that contains about $150k$ events and for each of them there are the following variables:

- Jets of the event, each one characterised by p_T, η, ϕ, mass and CSV;
- Muons p_T, η, ϕ, mass and iso03;
- Electrons p_T, η, ϕ and mass;
- Missing energy p_t and ϕ;
- Information at generation level of the two parton from the Higgs (p_T, η, ϕ and mass).

With the generation level (gen-lev) information is possible to find which jets correspond to those generated by the two b quark produced in the decay of the Higgs. At each parton from the Higgs the jet closer in angle is associated using as discriminant ΔR (Figure 2.1), defined as follows

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.1.1}$$

The hadronization of a b quark does not always produce a collimated jet and this behaviour may cause problems for the matching (large value of ΔR)

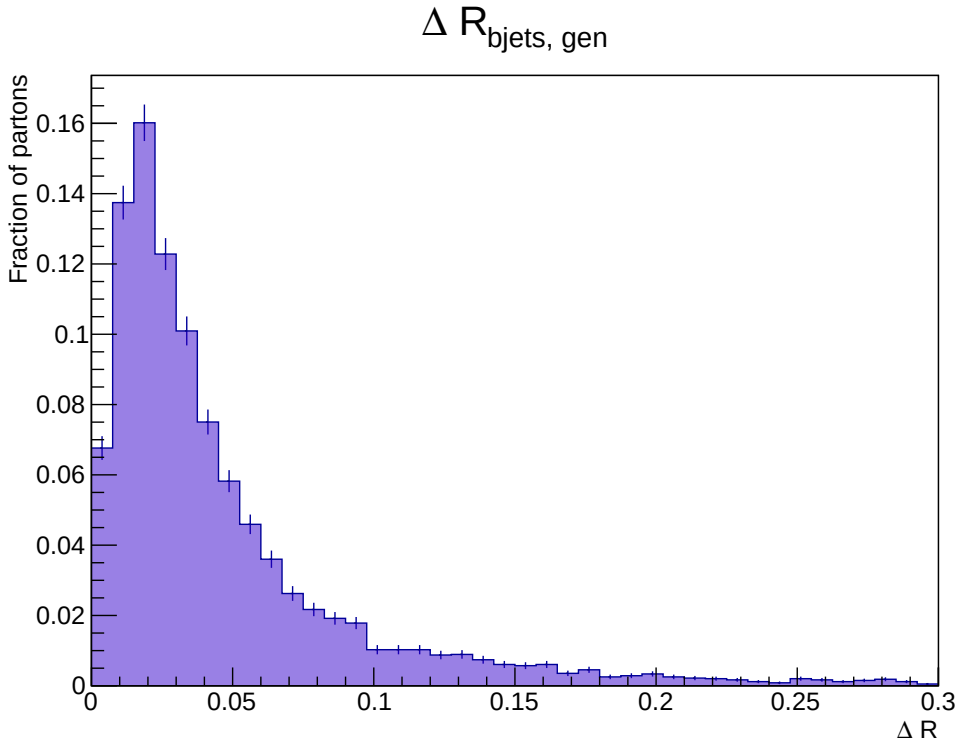


Figure 2.1: ΔR value relative to the best matches jet-parton.

2.2 Preliminary selection

The focus of this work is the $t\bar{t}H$ production in the all-hadronic final state. In order to consider only the fully hadronic final state events with at least 8 jets are required with a $p_T \geq$20 GeV and pseudorapidity $|\eta| \leq$2.4. The fraction of events selected with this request is 15.11±0.09% where the fraction is $f = \frac{N_{selected}}{N_{events}}$ and the uncertainty is the Wald approximation[12] to the binomial 68% confidence interval:

$$\sigma_f = \sqrt{\frac{f\,(1-f)}{N_{events}}} \tag{2.2.1}$$

As it is possible to see in Figure 1.9 in the final state there are 4 b-quark jets. In order to reduce the QCD background component as much as possible, four b-tagged jets must be required in the data selection. As mentioned before in section **1.1.4**

a jet is b-tagged if its CSV value is greater than a fixed value: *Loose* (≥ 0.5426), *Medium* (≥ 0.8484) or *Tight* (≥ 0.9535). The fractions of events selected requiring 4 b-jets with different CSV value are reported in Table 2.1.

CSV	Fraction of selected events
Loose	9.57±0.07%
Medium	2.79±0.04%
Tight	0.46±0.02%

Table 2.1: Fraction of selected events requiring a number of jets ≥ 8 and at least 4 b-jets.

From this point, a jet with CSV$\geq$0.8484 (medium working point) is considered a b-jet. It is advantageous to also apply a veto on identified leptons, in order to select the final state of our interest: if in the event there is a identified muon or an electron the event is discarded. Whit these cuts the fraction of selected events is 1.93±0.04%.

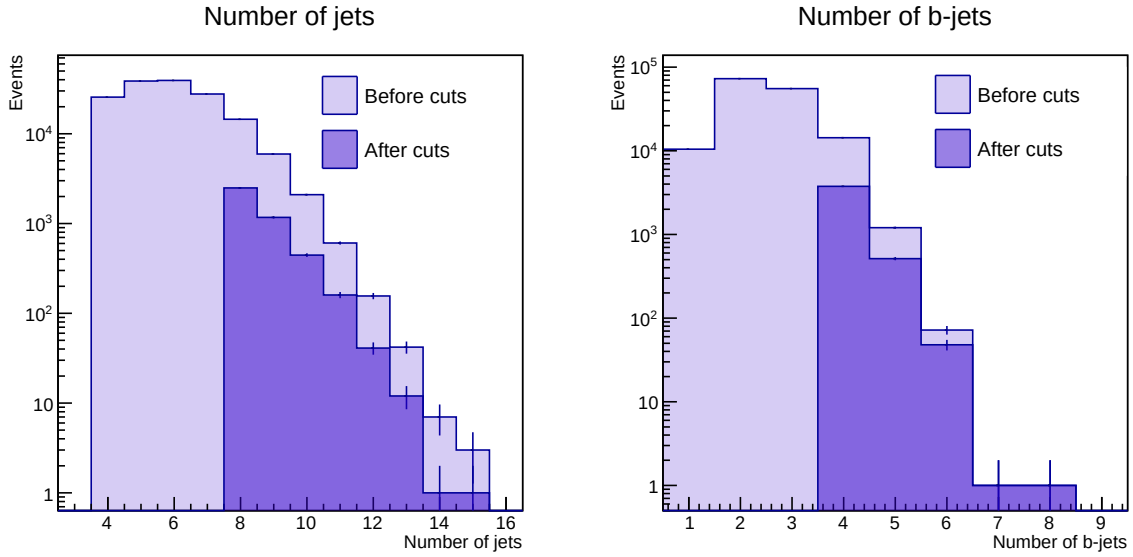


Figure 2.2: Number of jets and b-jets in the signal simulation before and after the cuts.

In Figure 2.2 it is possible to see the number of jets and b-jets before and after the cuts described previously. Because of the large number of jets and b-jets (respectively up to 15 and 8) it is not easy to assign jets to the final state parton therefore it is not easy to identify the two b-jets coming from the decay of the Higgs.

Chapter 3

Analysis and reconstruction of $t\bar{t}H$

In this chapter is described how the reconstruction of the system has been performed: the goal is to associate a jet to each final state parton. A precise identification of the jets coming from the final state partons of the $t\bar{t}H$ decay will allow the construction of discriminating features with which to later operate a classification of signal and background events, in view of a more stringent data selection. Due to the large number of jets in the final state this association is not trivial and it will lead to a combinatorial problem. A kinematic fit of the final state to $t\bar{t}H \to 8$ jets signal hypothesis will be performed using three different χ^2 statistics which are described in the following section.

3.1 Combinatorial problem

The final state can be reconstructed from the jets by using a kinematic fit that associates each jet to the final state parton which originated it. In the final state, as discussed before, there are 8 jets: there are 4 jets coming from light quark j_1, j_2, j_3, j_4 and 4 b-jets bj_1, bj_2, bj_3, bj_4. Because of the fact that there can be more than 8 jets, the 4 with higher CSV are used as b-jets and the 4 (or 5, as we will see later) with higher p_T as the light quark products. Due to the large number of jets the problem of combinatorics arises: with a simple example we can easily see the large number of possible combinations that affect this channel. Suppose that $W^+ \to j_1 j_2$ then $t \to W^+ b j_1$ and in a similar way $W^- \to j_3 j_4$ then $\bar{t} \to W^- b j_2$ thus the two remaining b-jets are the ones from the Higgs $H \to b j_3 b j_4$. Fixing the b-jets it is possible now to swap the jets in 6 different way and then, in a similar way, swap also the b-jets in 6 different ways maintaining the position of jets fixed. For each permutation of the b-jets there are 6 possible permutations of the jets. Therefore there are 6·6=36 possible combinations. More in general, indicating with N_{jets} the

number of jets and N_{b-jets} the number of b-jets there are

$$N = \binom{N_{jets}}{2} \cdot \binom{N_{b-jets}}{2} \tag{3.1.1}$$

combinations. Thus, if we use 5 jets instead of 4 and 4 b-jets there are N=60 possible combinations.
The large number of possible combinations make the identification of the correct topology difficult, but it is possible to use some constrains like the mass of the quark top m_t, $W^{\pm}$ boson m_W and even the Higgs mass m_H. If we do not use the mass of the top (or the mass of the Higgs) as constraints it is possible to use it after the reconstruction as a control and also it will not introduce a bias in the background thus it is possible to use them as discriminant between signal and background at a later stage.

3.2 Kinematic fit

One possible approach that can be used to reconstruct the system is the χ^2: For each possible combination of jets and b-jets described in the previous section a χ^2 can be calculated; then, the best combination is the one with the smallest χ^2. A request of Prob(χ^2) $\geq$1% can be applied in order to strongly reduce the background and the wrong events selection. The value 1% is not strictly relevant here: The aim of this chapter is to test the relative power of the three different χ^2, so 1% is taken as a reference to consider only events that can be reconstructed reasonably well. The kinematic fit of the system will be performed in three different ways, indeed it is possible to build three different types of χ^2 by including more information about the physical constraints coming from the mass of the decaying particles.

3.2.1 χ^2_1 with three constraints

The first χ^2 can be built to constrain the $W^{\pm}$ masses and requiring the equality of the top masses

$$\chi^2_1 = \frac{(m_{W^+} - m^*_W)^2}{\sigma^2_{m_W}} + \frac{(m_{W^-} - m^*_W)^2}{\sigma^2_{m_W}} + \frac{(m_t - m_{\bar{t}})^2}{\sigma^2_{\Delta m_t}} \tag{3.2.1}$$

were m^*_W=80.4 GeV/c^2, σ_{m_W}=15 GeV and σ_{m_t}=30 GeV then $\sigma_{\Delta m_t} = \sqrt{\sigma^2_{m_t} + \sigma^2_{m_{\bar{t}}}}$ $= \sqrt{2}\sigma_{m_t} \approx 42$ GeV. These σ values are equal to typical experimental resolution of masses reconstructed by the decay in two or three jets. The two remaining b-jest are considered as coming from the decay of the Higgs.

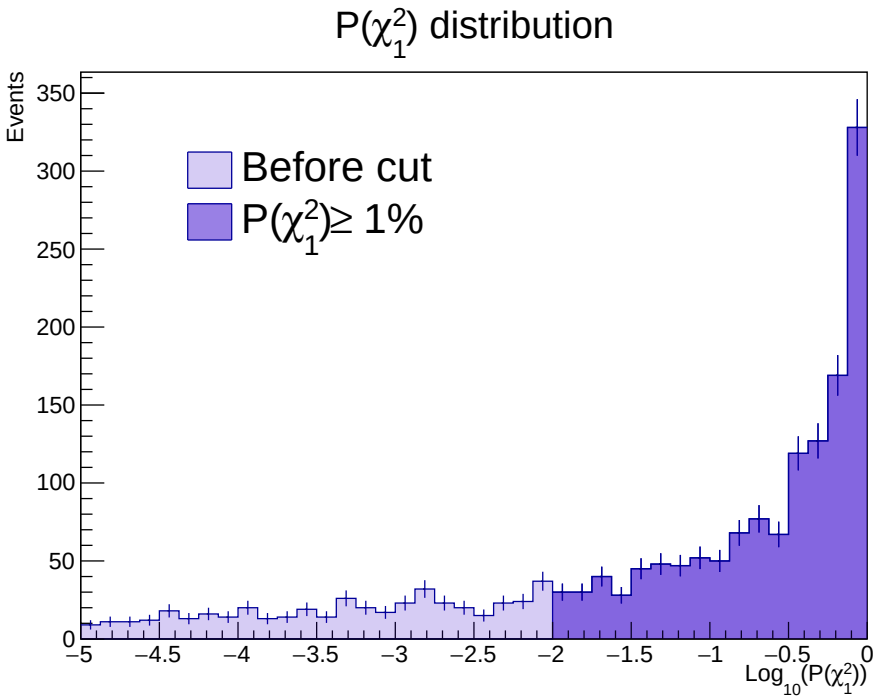


Figure 3.1: Distribution of the best χ^2_1 before ad after requiring Prob(χ^2_1)≥1%.

As has been said before (section 3.1), for each event this χ^2_1 has been computed using the 4 b-jets with higher CSV and the 4 jets with higher p_T and then, if Prob(χ^2)≥1% the event is accepted. In Fig 3.1 it is possible to see the distribution of the best χ^2_1 before and after the request of Prob(χ^2_1)≥1%.
The fraction of selected events is $43.7 \pm 0.9\%$ thus the fraction of events that has not been selected is $56.3 \pm 0.9\%$.
As it is possible to see in Fig. 3.2 it is generally easy to find a combination of jets compatible with the W mass and, even without constraining the top quark mass, the minimum χ^2 requirement correctly identifies the decay of that particles (Fig. 3.3).
Using the generator information, as described in section 2.1, it is possible to determine which b-jets are closer (in term of ΔR) to the two b-jets coming from the Higgs and then compare them with the two b-jets found with the χ^2_1. In table 3.1 are reported the fractions of events where there are 2 matches, 1 and 0 matches: 2 matches means that the two b-jets identified as the two produced by the Higgs with χ^2_1 correspond to the ones found by using the generator level information.
In the table there are 2 rows: the first one refers to the selected events (the ones with Prob(χ^2_1)≥1%) and the second to the ones with Prob(χ^2_1)<1%.

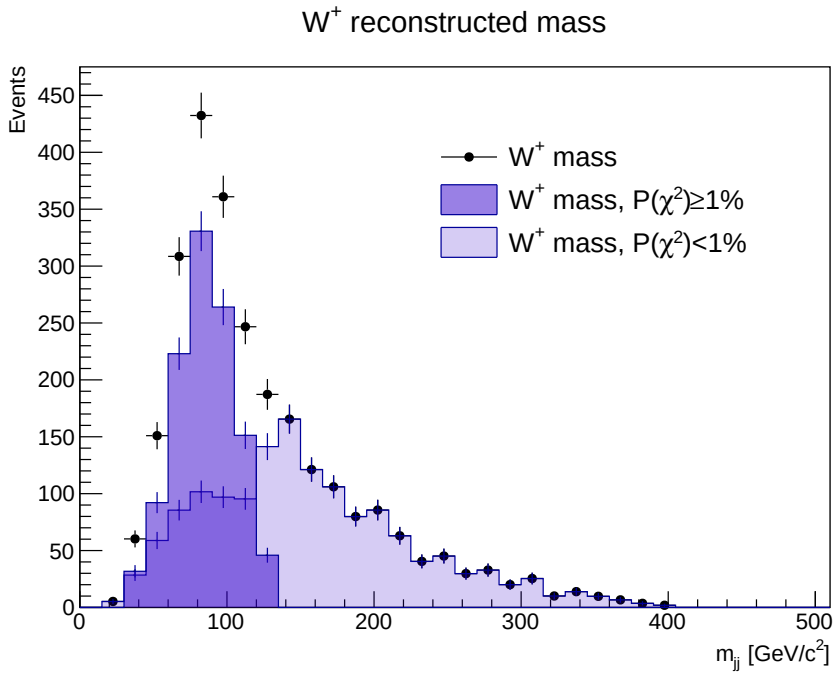


Figure 3.2: Reconstructed W mass before and after requiring Prob(χ_1^2)$\geq$1%.

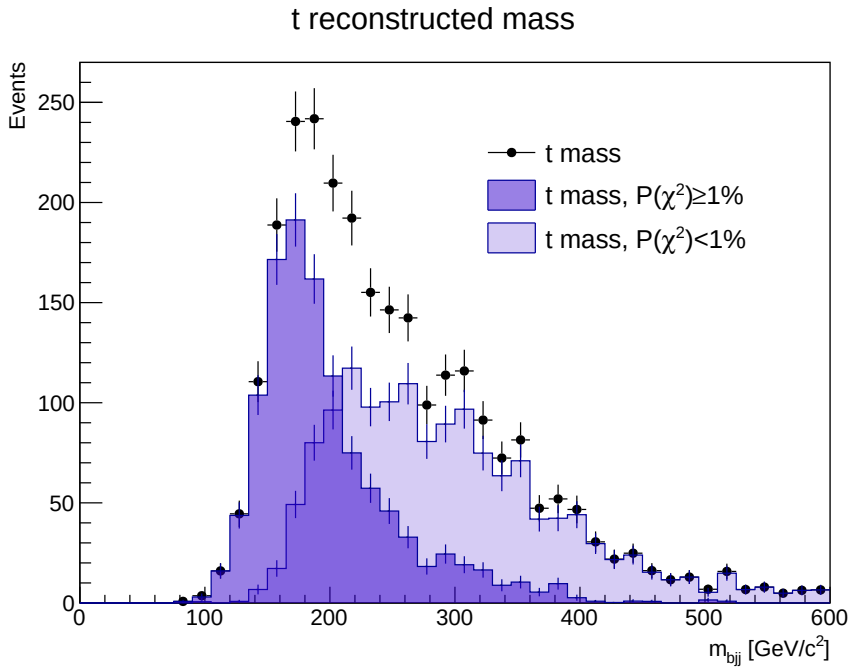


Figure 3.3: Reconstructed t mass before and after requiring Prob(χ_1^2)$\geq$1%.

	2 matches (%)	1 match (%)	0 match (%)
P(χ^2_1)≥1%	22±1	64±1	15± 1
P(χ^2_1)<1%	15±1	67±1	17± 1

Table 3.1: Fraction of events with 2,1,0 matches of the 2 b-jets found with the χ^2_1 and the generator level ones.

In Fig. 3.4 it is possible to see the Higgs reconstructed mass for the selected events and for the events with 2,1,0 matches. As expected the Higgs mass is well reconstructed for the events with 2 matches.

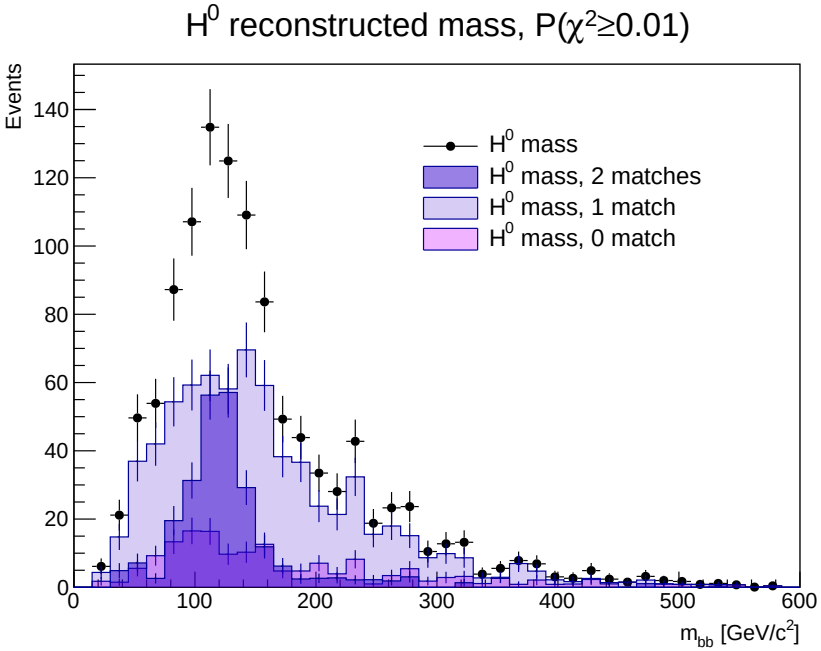


Figure 3.4: Higgs reconstructed mass for the selected events (Prob(χ^2_1)≥1%) and the number of matches.

The events discarded before (the one with Prob(χ^2_1)<1%) have been reanalysed in the same way as before but using 9 jets instead of 8: 5 jets from light quark and 4 b-jets. In the class of events previously discarded (56.3±0.9%) the fraction which after the reanalysis has Prob(χ^2_1)>1% is 18.8±0.7% and the fraction of events discarded is 37.4±0.9%.
It is now possible to combine the two classes: if the first analysis discards one event which has at least 9 jets, it will be reanalysed considering 9 jets in the combinatorial problem and accepted if the best combination give a χ^2_1 which has a probability

Prob(χ^2_1)≥1%.
The fraction of selected events is now 62.6±0.9% (43.7±0.9% before) and the fraction of discarded events is 37.4±0.9%(56.3±0.9% before).
In figure 3.5 it is possible to see the reconstructed mass of the top quark using 9 jets for the selected and discarded events compared with the reconstructed mass with 8 jets (see figure 3.3). In table 3.2 it is possible to see the new fraction of events with 2,1,0 matches. The fraction of events with 2 matches decreases by 1%.

	2 matches (%)	1 match (%)	0 match (%)
P(χ^2_1)≥1%	20.4±0.9	64±1	15.2±0.8
P(χ^2_1)<1%	17±1	66±1	17± 1

Table 3.2: Fraction of events with 2,1,0 matches of the 2 b-jets found with the χ^2_1 and the generator level ones for the reanalysis.

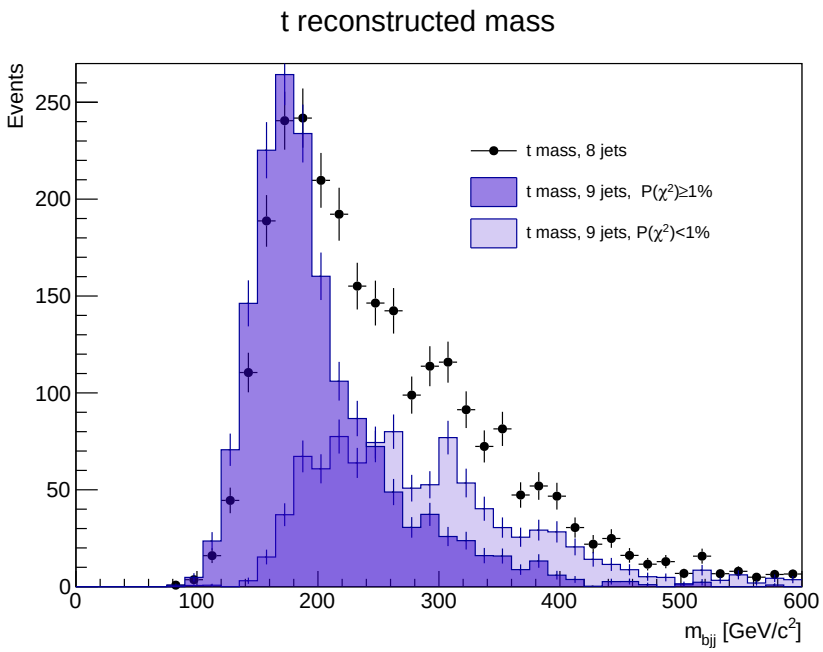


Figure 3.5: Reconstructed t mass before and after requiring Prob(χ^2_1)≥1% using 9 jets. The black dots represent the mass reconstructed using 8 jets.

In figure 3.6 is shown the reconstructed mass of the t quark vs. $\bar{t}$ quark. As it is possible to see the mass of the two top quarks is well reconstructed, this means that the used χ^2_1 seems to identify easily triplet of jets compatible with the top quark mass even without constraining it.

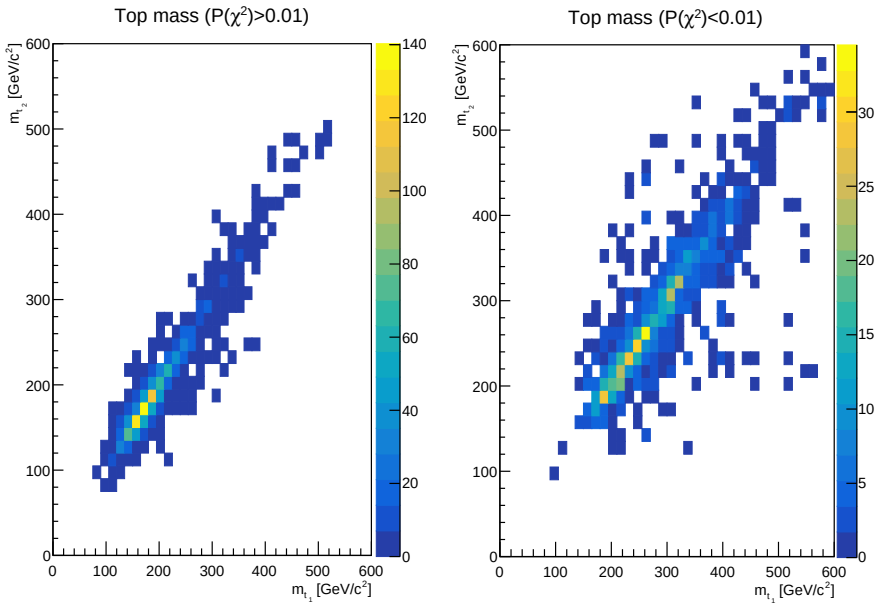


Figure 3.6: t vs $\bar{t}$ reconstructed mass for the selected and discarded events.

3.2.2 χ^2_2 with four constraints

As has been done before for the χ^2_1, it is possible to build χ^2_2 constraining also the top quark masses to m_t=173.3 GeV instead of constraining only the W boson masses

$$\chi^2_2 = \frac{(m_{W^+} - m^*_W)^2}{\sigma^2_{m_W}} + \frac{(m_{W^-} - m^*_W)^2}{\sigma^2_{m_W}} + \frac{(m_t - m^*_t)^2}{\sigma^2_{m_t}} + \frac{(m_{\bar{t}} - m^*_t)^2}{\sigma^2_{m_t}} \tag{3.2.2}$$

where σ_{m_W}, σ_{m_t} are the ones used before. As it is possible to see in figure 3.7 the distribution of χ^2_2 is similar to the one of χ^2_1.

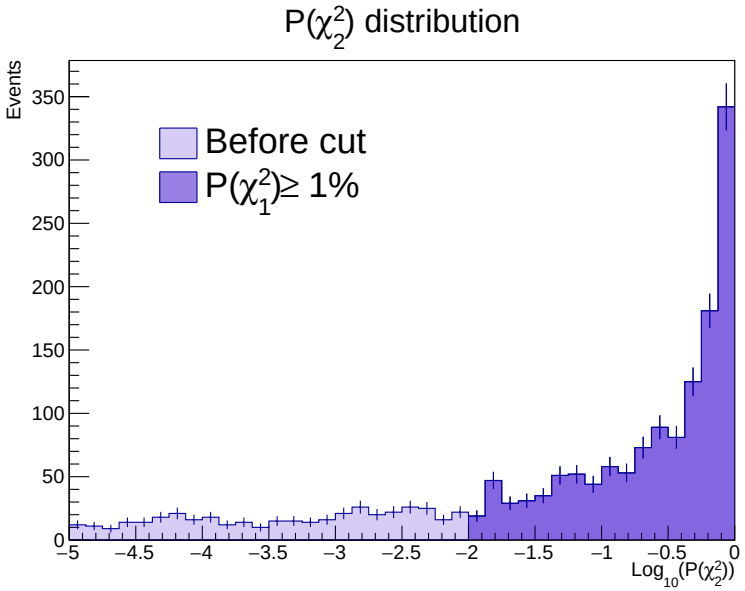
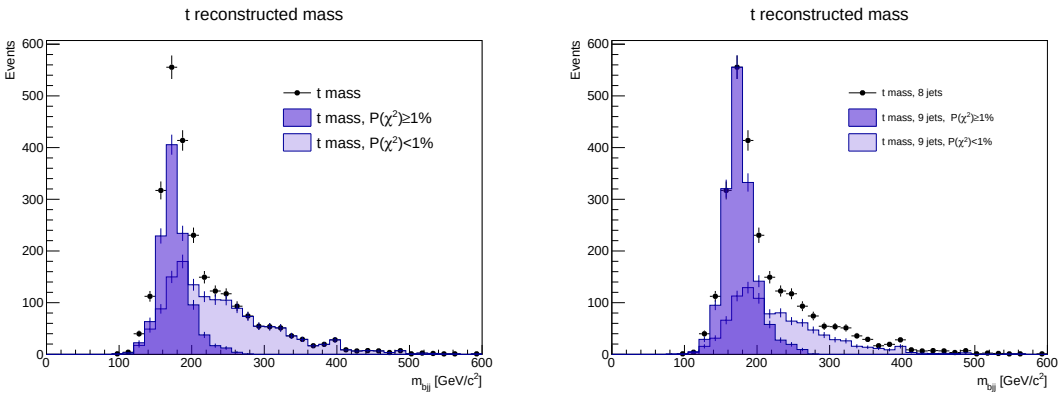


Figure 3.7: Distribution of the best χ^2_2 before ad after requiring Prob(χ^2_2)≥1%.

Again, if an event has P(χ^2_2)<1% and at least 9 jets il was reanalysed in the same way as described before. The fraction of selected events is 62.1±0.9% and the discarded one is 37.9±0.9%.
In figure 3.8(a) is reported the top quark mass reconstructed using 8 jets and in figure 3.8(b) the one reconstructed using 9 jets (The black dots represent the reconstructed mass using 8 jets). As one expected, constraining also the mass of the top quark improves a lot the quality of the reconstruction.
The fraction of events with 2,1,0 matches are reported in table 3.3. The first row refers to the selected events considering only 8 jets, on the other hand the second row refers to the events selected using 9 jets. With respect to what happens for the χ^2_1 the fraction of events with two matches increases by 4% if we consider the reconstruction with 8 jets or 3% for the reconstruction with 9 jets. The reconstructed Higgs mass is shown in figure 3.9.



(a) (b)

Figure 3.8: Reconstructed mass of the top quark using 8 (a) and 9 jets (b).

	2 matches (%)	1 match (%)	0 match (%)
$P(\chi^2_2)\geq 1\%$, 8 jets	24±1	64±1	12±1
$P(\chi^2_2)\geq 1\%$, 9 jets	23±1	64±1	13± 1

Table 3.3: Fraction of events with 2,1,0 matches of the 2 b-jets found with the χ^2_2 and the generator level ones for the reanalysis.

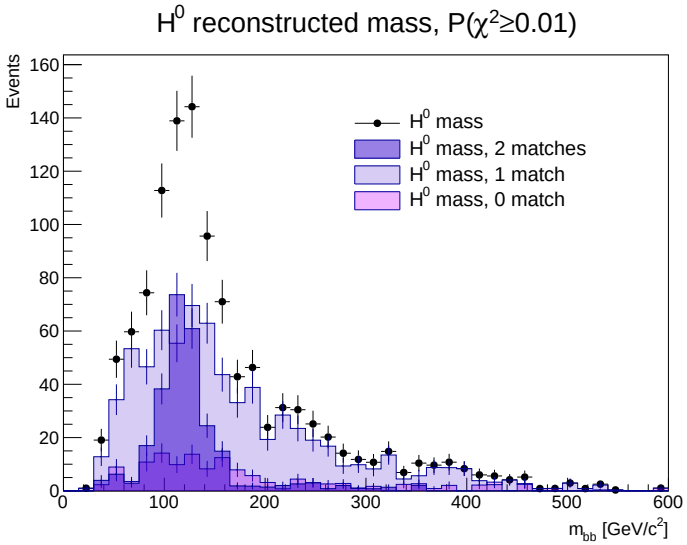


Figure 3.9: Higgs reconstructed mass for the selected events (Prob(χ^2_2)$\geq$1%) and the number of matches.

3.2.3 χ^2_3 with five constraints

Like what has been done for the χ^2_1 and χ^2_2 it is possible tu build a third χ^2_3 constraining also the mass of the Higgs at the value of 125 GeV

$$\chi^2_3 = \frac{(m_{W^+} - m^*_W)^2}{\sigma^2_{m_W}} + \frac{(m_{W^-} - m^*_W)^2}{\sigma^2_{m_W}} + \frac{(m_t - m^*_t)^2}{\sigma^2_{m_t}} + \frac{(m_{\bar{t}} - m^*_t)^2}{\sigma^2_{m_t}} + \frac{(m_H - m^*_H)^2}{\sigma_{m_H}} \tag{3.2.3}$$

where σ_{M_H}=20 GeV. The distribution of the χ^2_3 is similar to the previous ones (Fig. 3.10).

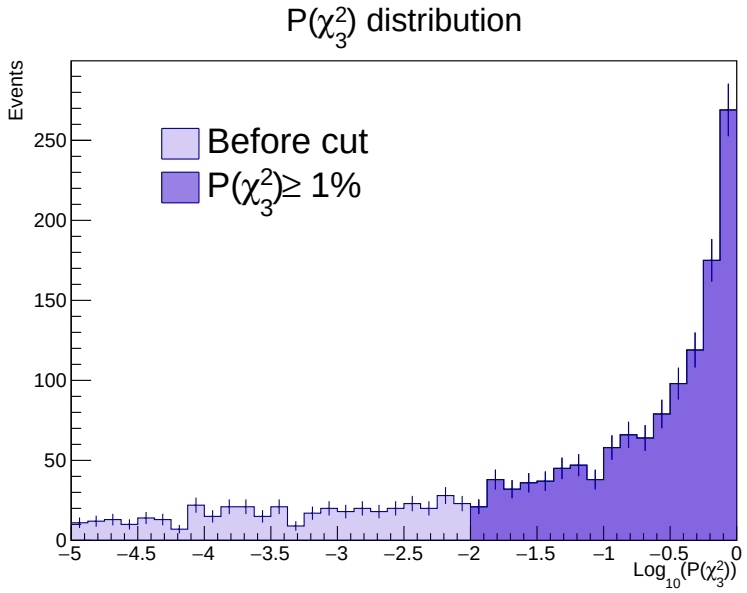
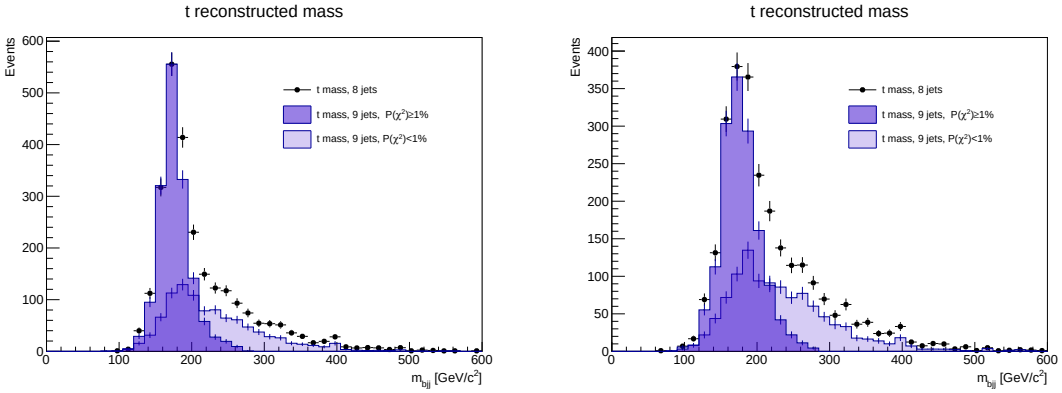


Figure 3.10: Distribution of the best χ^2_3 before ad after requiring Prob(χ^2_3)≥1%.

Again, if an event has P(χ^2_3)<% and at least 9 jets it was reanalysed in the same way as described for the χ^2_1. The fraction of selected events is 54.5±0.9% and the discarded one is 42.5±0.9% and, as expected, this is lower than the one obtained with χ^2_1 (62.6±0.9%).
In figure 3.11 are shown the reconstructed masses of the top quark using χ^2_3 (Fig. 3.11(a)) and χ^2_3 (Fig. 3.11(b)). Despite the fact that using χ^2_3 the mass of the top is reconstructed correctly, the distribution result wider than the one obtained using χ^2_2. This is due to the fact that χ^2_3 has smaller constraining power to find the combination of jet triplets closer to 173.3 GeV. Once a MC simulation where a full information of the generator-level information of W and top decay products will be available, it will be possible to directly compare the fraction fraction of combinations that correctly identify the jets originated by the W and top decays, as we could do here only in the case of the H→b$\bar{\text{b}}$ decay products.

The fraction of events with with 2,1,0 matches are reported in table 3.4. The first row of the table refers to the selected events using 8 jets, the second one to the selected events using 9 jets. With respect to what happens with the previous χ^2_1 and χ^2_2 the fraction of events with 2 matches is now almost doubled. Again it is possible to notice that the fraction of events with 2 matches decrease for the reanalysed events by 2%. In figure 3.12 is shown the mass distribution of the Higgs boson for the reconstructed events with 2,1,0 matches.



(a) (b)

Figure 3.11: Reconstructed mass of the top quark using χ^2_2 with 9 jets(a) χ^2_3 with 9 jets(b).

	2 matches (%)	1 match (%)	0 match(%)
P(χ^2_3)≥1%, 8 jets	40±1	49±1	10.8±0.9
P(χ^2_3)≥1%, 9 jets	38±1	50±1	11.5±0.8

Table 3.4: Fraction of events with 2,1,0 matches of the 2 b-jets found with the χ^2_3 and the generator level ones for the reanalysis.

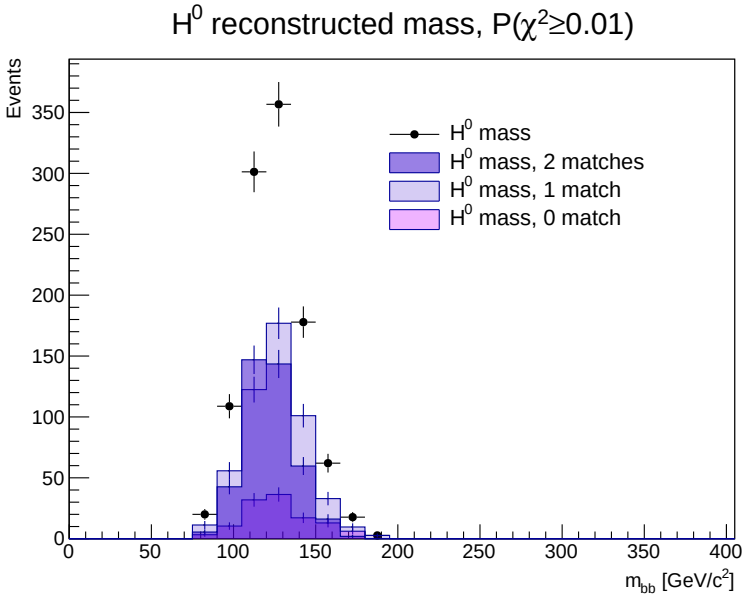


Figure 3.12: Higgs reconstructed mass for the selected events (Prob(χ^2_3)$\geq$1%) and the number of matches.

Chapter 4

Conclusions

In this work three different ways to reconstruct the final state of fully hadronic $t\bar{t}H$ decays were presented. After an initial preselection of the events, needed in order to discard leptonic and semi-leptonic events, the final state has been reconstructed through a kinematic fit. The first out of three types of χ^2 was built constraining only the W bosons masses and requiring the equality of the reconstructed masses of the top quarks. This χ^2_1 has been computed for each possible combination of 8 jets (with 4 b-jets) and the one with the best χ^2_1 has been selected as best combination. If for the best combination Prob(χ^2_1)$\geq$1% the event passed the selection. If an event has Prob(χ^2_1)<1% and at least 9 jets it was reanalysed considering the possible combinations with 9 jets (4 b-jets) and if Prob(χ^2_1)$\geq$1% the event passed the selection. Thanks to the generation-level information of the b quarks emitted in the Higgs decay it was possible to check if the two b quarks identified as the ones generated from the decay of the Higgs boson correspond to the two of generation level. For the χ^2_1 the fraction of events with 2 matches is 20.4$\pm$0.9%.
A second χ^2_2 was built constraining also the masses of the two top quarks. Proceeding as described before the fraction of events with two matches becomes 23$\pm$1%. The last χ^2_3 was built constraining the masses of W bosons, top quarks and Higgs boson. The fraction of events with two matches becomes now 38$\pm$1%.
The ability to decrypt the complex topology assigning jets to final state partons with a reasonably good fraction of matches is a tool that can be used to study the kinematics of the signal.
Further studies can be done once the all hadronic $t\bar{t}$ decays will be preselected at generator level and this can be compared with the background in order to construct a multivariate discriminant.

List of Figures

List of Tables

Bibliography

[1] CMS collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716 (Sept. 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[2] G. Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716 (Sept. 2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[3] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[4] Albert M Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: (2017). arXiv: 1706.04965 [physics.ins-det].

[5] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Anti-k(t) jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[6] Serguei Chatrchyan et al. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[7] Georges Aad et al. "Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].

[8] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[9] CMS Collaboration. "Search for $\mathrm{t\bar{t}}H$ production in the $\mathrm{H}\to\mathrm{b\bar{b}}$ decay channel with 2016 pp collisions data at $\sqrt{s}$=13 TeV". In: (2016). URL: https://cds.cern.ch/record/2231510/files/HIG-16-038-pas.pdf.

[10] C. Oleari. "The Powheg Box". In: *Nuclear Physics B Proceedings Supplements* 205 (Aug. 2010), pp. 36–41. DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph].

[11] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[12] R. D. Cousins, K. E. Hymes, and J. Tucker. "Frequentist evaluation of intervals estimated for a binomial parameter and for the ratio of Poisson means". In: *Nuclear Instruments and Methods in Physics Research A* 612 (Jan. 2010), pp. 388–398. DOI: 10.1016/j.nima.2009.10.156. arXiv: 0905.3831 [physics.data-an].