University of Zurich

Faculty of Mathematics and Physics

**BACHELOR THESIS**

Fernando Moreno Fernández

# Developments in the search for 4 top production at CMS

Particle Physics Department

| | |
|---|---|
| Supervisor of the bachelor thesis: | Florencia Canelli |
| Study program: | Bachelor in Physics |
| Specialization: | Applied Physics |

Zurich 2022

**Acknowledgments:** First of all I would like to thank Florencia Canelli for agreeing to supervise my thesis and giving me the opportunity to work in her group. On the other hand I am eternally grateful to Sergio Sánchez Cruz for all his help and patience, without him this thesis would not have been possible. Finally, I would like to thank all my family who have always supported me and all the friends I have made over the years, especially : Alain, Sergio, Fran, Hugo, Duje and Miguel, without your friendship and help I would never have reached this point.

**Abstract:**

The production of $t\bar{t}t\bar{t}$ in $pp$ collisions has been studied in CMS at $\sqrt{s} = 13$ TeV. We develop an analysis aimed to search for $t\bar{t}t\bar{t}$ production. The analysis developed uses regions with two same-sign and three leptons in the final state. We optimise the selection based on the number of b-tagged jets and the total transverse momentum. The sensitivity of the designed analysis is estimated using simulations, achieving an expected significance of 2.8 standard deviations.

**Keywords:** top quark, four top quarks, multileptons, CMS, LHC

# Contents

# Introduction

The aim of this thesis is to design an analysis to study the production of $t\bar{t}t\bar{t}$ in pp collisions at the LHC with the largest possible sensitivity. For this purpose, Monte Carlo simulations that replicate the conditions of the detector in 2018 will be studied. Although the simulations correspond to the year 2018, a luminosity of 140 $\text{fb}^{-1}$ has been used, which corresponds to the entire Run 2 (2016, 2017 and 2018). It is assumed that $t\bar{t}t\bar{t}$ exists as described in the Standard Model (SM) and that the results of the simulations agree with what would be observed in CMS. It should be noted that the $t\bar{t}t\bar{t}$ process has been studied previously as can be seen in [9], unlike previous work, this thesis features improvements in lepton reconstruction.

In section 1, $t\bar{t}t\bar{t}$ is framed within the SM and the LHC experiment, more specifically in the CMS detector. Section 2 describes the $t\bar{t}t\bar{t}$ process and the main backgrounds that will appear during the analysis. Then, in section 3, the analysis is described in detail: what types and how the regions are constructed, the lepton selection used, and the event selection criteria used. Section 4 shows all the contributions of each process to the different regions, calculates the signal strength, computes the sensitivity of the analysis and presents different sources of uncertainty. And finally, the last section shows the final conclusions of this work.

# 1. The SM and $t\bar{t}t\bar{t}$ production at the LHC

## 1.1 The Standard Model

The universe is composed of energy and matter. At the most fundamental level it is formed by elementary particles. Currently the interactions between the different elementary particles are described by a quantum field theory called the Standard Model (SM).
The SM is composed of two different types of particles: fermions and bosons. Fermions have half-integer spin, and compose matter in the universe. On the other hand, bosons (also called force carriers) have integer spin and carry the interactions among the different elementary particles [8].
Fermions are divided in two groups depending whether they are subject to the strong interaction or not, they can be quarks (up, down, top, bottom, strange and charm), which are subject to the strong interaction, or leptons (electron, muon, tau and their three corresponding neutrinos), which are not. The bosons are: the $Z$ and $W^{\pm}$ bosons and the photon, which are the mediators of the electroweak interaction, the gluons, which are involved in the strong interaction and finally the Higgs boson. [8].

All SM particles and their characteristics are shown in Figure 1.1.

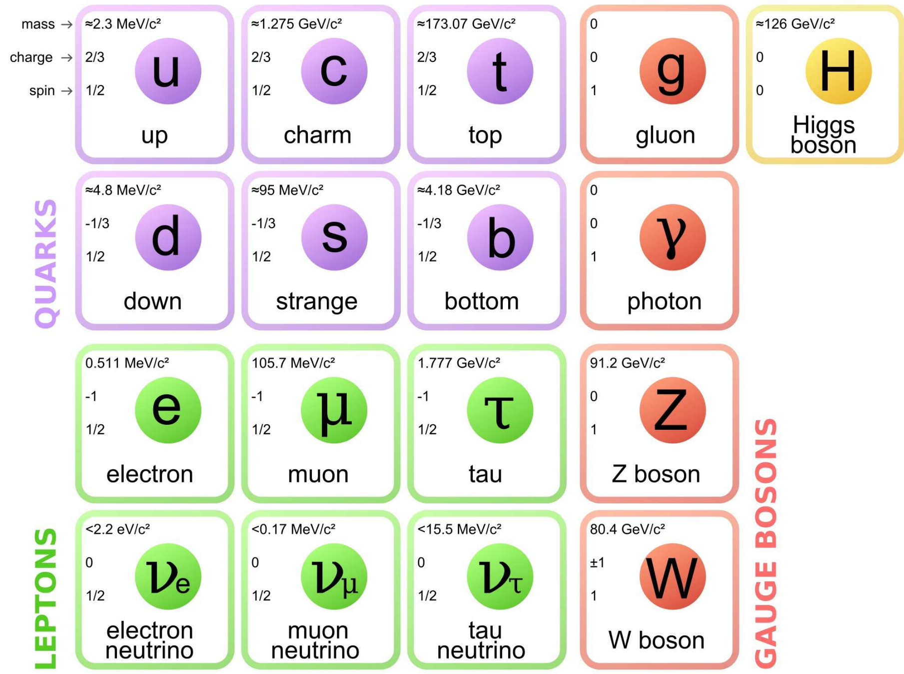


Figure 1.1: All currently known SM particles and their properties are shown. Image taken from [5]

Currently the SM is not considered to be a complete theory as it is not able to describe several phenomena, among them the gravitational force (a gravitational force carrier has not been discovered) or the presence of dark matter. Current evidence points to the need to go beyond the standard model (BSM), dark matter or oscillations in neutrino mass are some of the main objects of study in BSM physics.

## 1.2 Beyond SM and the importance of $t\bar{t}t\bar{t}$

The $t\bar{t}t\bar{t}$ process is quite rare within the SM , with a predicted cross section in the SM of $12.0 \pm 2.4$ fb at 13 TeV as is calculated in [9]. It is always interesting in itself to study a rare process in the SM, however BSM physics could affect the production of $t\bar{t}t\bar{t}$, and because some BSM models affect $t\bar{t}t\bar{t}$ production, studying $t\bar{t}t\bar{t}$ is specially interesting to provide clues about which direction BSM physics should go.
The contribution of BSM to $t\bar{t}t\bar{t}$ production can be mainly divided into two types: new BSM interactions (possibly due to high-mass particles) and on-shell production of BSM particles. [10].

## 1.3 LHC and CMS

To produce particles with very high mass, such as top quarks, we need to collide particles at a very high energy, for which particle accelerators are built. Beams of protons (although also heavy ions) are accelerated and made to collide at very high energy. Right at the point of collision, detectors are built to study all the products of the collision. This work is performed considering proton-proton collisions produced in the CERN Large Hadron Collider (LHC) and recorded by the CMS detector.
The LHC is a particle accelerator of 27 km in circumference, located mainly in Switzerland, it archives the highest centre of mass energy (up to $\sqrt{s} = 13.6$ TeV) to date, which, together with its high luminosity, makes it the most powerful accelerator in the world.
The LHC is not really a perfect circle, but is composed of arcs and inserts (see figure 1.2), the arcs contain magnetic dipoles, used for curving the trajectory of the proton beam so it keeps doing circles in the accelerator, and auxiliary magnets which focus the beam and direct it to the collision point. The inserts have different purposes such us: collisions in an experiment, injecting or discharging proton beams [12].
A proton source (currently Linac4) is used to generate the proton beams, the proton beams are injected into auxiliary accelerators which then inject the partially accelerated beam into the main ring of the LHC where they are accelerated to near-light speed and collide at one of the detector sites [12].
One of the main detectors of the LHC is the Compact Muon Solenoid (CMS), a general detector weighing 14,000 tonnes and measuring only 15 metres high and 21 metres tall [7], being quite **compact** compared to other detectors such as ATLAS. It characterised by its high efficiency in detecting **muons**, particles found in the final state of many processes of interest ($t\bar{t}t\bar{t}$ for example), with a high resolution in both the measurements of the momentum of charged particles up to energies of hundreds of GeV and the measurements of the missing transverse momentum. It is built around a large **solenoid** which generates a magnetic field

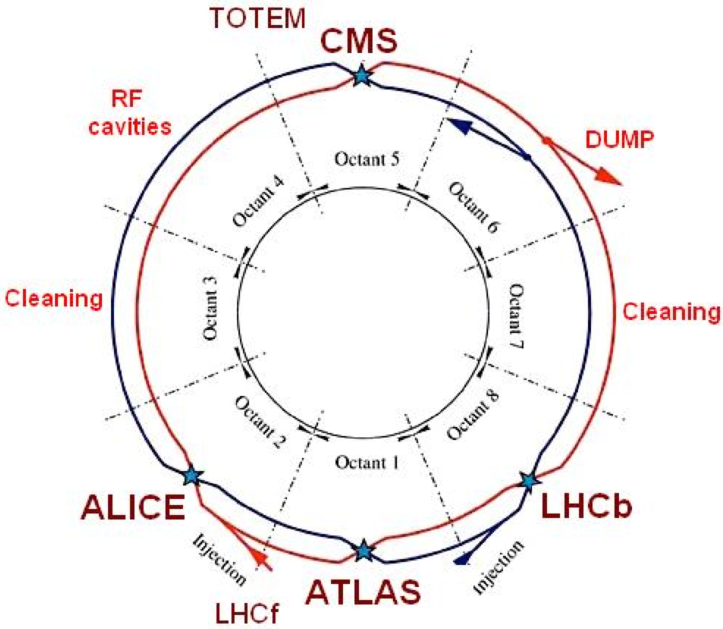


Figure 1.2: A schematic of the LHC, its sections, inserts and main detectors are shown. Figure taken from [12]

of 3.8 T, used to bend the trajectory of particles produced in the detector, this allows to measure their momentum in the transverse direction. The main parts of CMS are shown in figure 1.3.

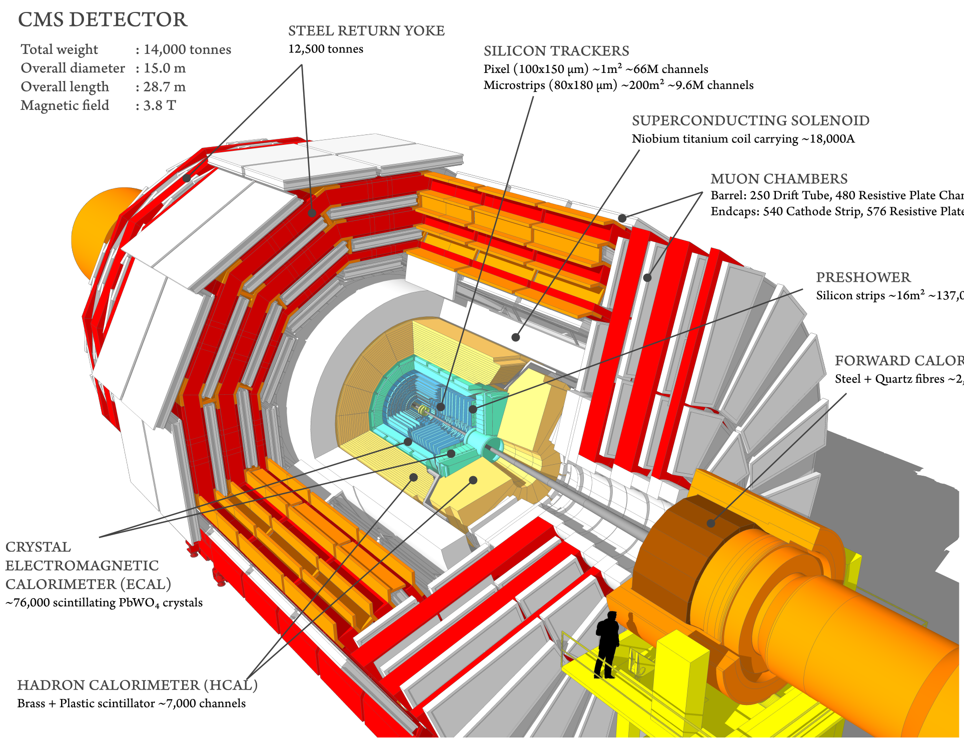


Figure 1.3: A schematic of the different parts and the number of components of each part of CMS is shown, figure taken from [7]

# 2. Analysis Overview

## 2.1 The signal $t\bar{t}t\bar{t}$

As mentioned before, the $t\bar{t}t\bar{t}$ process has a very small cross section within the SM ($12.0 \pm 2.4$ fb at 13 TeV, [9]), much smaller than the main backgrounds expected in the analysis such us $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$, so it is important to know what particles are produced in the process in order to use variables that can be useful in distinguishing the signal from the background during the analysis.

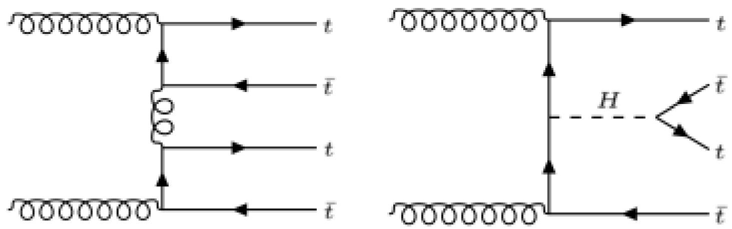


Figure 2.1: Two Feyman diagrams of the $t\bar{t}t\bar{t}$ process are shown, image taken from [9]

Top quarks are quite heavy and therefore decay into lighter particles, which are measured in the detectors. In particular, the top quark decays into a bottom quark and a $W$ boson. In turn, the $W$ boson can decay into a lepton and a neutrino (in about 10% of the cases for each flavor of leptons) or into a pair of quarks ($u$, $c$, $d$, $s$) in the other cases. After the collision at the interaction point and before reaching the detector, particles that have color charge irradiate through strong interaction other particles, this spray of particles is known as a jet. In this case the particles producing the jets are the bottom quarks so they are called b-jets.

Therefore, after the decay of the four top quarks, one expects to find b-jets and between 0 and 4 leptons with their corresponding neutrinos. As it is not possible to directly detect the neutrinos in CMS, it is more convenient to focus on the presence of leptons and b-jets. One distinctive feature of the $t\bar{t}t\bar{t}$ process is that it gives rise to events containing a pair of same-sign leptons (2lss), since they can come from the decay of two same-sign top quarks. This signature is quite rare in the Standard Model, since particles are usually produced in pairs of particle-antiparticle. Therefore it is interesting to focus on the presence of 2lss in the final state to reduce lot of possible backgrounds.

## 2.2 Backgrounds in the Analysis

There are processes that also give rise b-jets and 2lss in the final state, for that reason they will be the main backgrounds in this analysis.

The most relevant ones are $t\bar{t}Z$, $t\bar{t}W$, $t\bar{t}H$ and processes with non-prompt leptons.

It is important to study also the backgrounds and not only the signal, this way the analysis can be optimized to reduce the amount of background. In addition, understanding the backgrounds and how they may affect the analysis allows for better estimation of uncertainties and awareness of the limitations of the analysis.

### 2.2.1 $t\bar{t}W$

The $t\bar{t}W$ process will be one of the main backgrounds contributing to the four top search because it can give raise to 2lss or 3l final state: each of the two top quarks decays into a $W$ boson and a b-quark, giving a total of three $W$ bosons, so two must necessarily have the same sign. More jet and b-jets are expected in $t\bar{t}t\bar{t}$ than in $t\bar{t}W$, however since the cross section of $t\bar{t}W$ ($\sigma_{t\bar{t}W} = 870 \pm 130_{stat} \pm 140_{syst}$fb measured at [1]) is much larger than the $t\bar{t}t\bar{t}$ one, it is expected to find much more background than signal in the regions defined in the analysis.
To mitigate this effect, regions with a large number of b-jets and with a large number of final state objects should be included in the analysis, as $t\bar{t}t\bar{t}$ generally produces more b-jets and a larger amount of final state objects than $t\bar{t}W$ simply due to the fact that the initial state of $t\bar{t}t\bar{t}$ has more particles capable of producing b-jets and more mass in total.

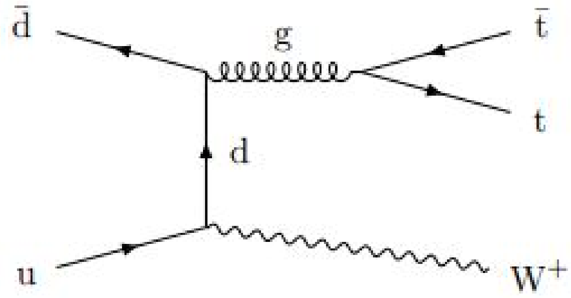


Figure 2.2: A representative Feyman LO diagram for $t\bar{t}W$ is shown. Image taken from [3]

### 2.2.2 $t\bar{t}Z$

$Z$ bosons can decay into a lepton-antilepton pair, a neutrino-antineutrino pair or a quark-antiquark pair. This, together with the $t\bar{t}$ pair, generates leptons which can be 2lss (if one of the leptons produced by the $Z$ boson is not correctly detected) or 3l and b-jets in the final state. Again, the cross section of $t\bar{t}Z$ ($\sigma_{t\bar{t}Z} = 950 \pm 80_{stat} \pm 100_{syst}$fb measured in [1]) is much larger than the $t\bar{t}t\bar{t}$ one. As with $t\bar{t}W$, adding regions with high b-jets and large number of final state objects can help to increase the number of signal events, as $t\bar{t}Z$ generally produces fewer b-jets and less number of final state objects.

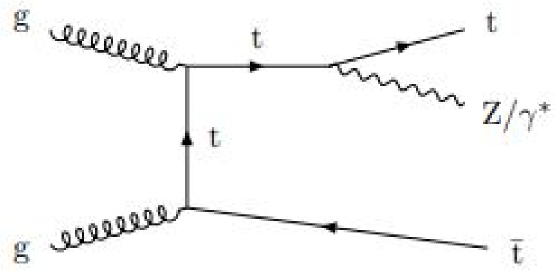


Figure 2.3: A representative Feyman LO diagram for $t\bar{t}Z$ is shown. Image taken from [3]

### 2.2.3 $t\bar{t}H$

The Higgs boson can decay in multiple ways, some of which give place to lepton pairs in the final state. For this analysis the most relevant decays are:
-The $H$ boson decays to a $W$ boson pair, decaying to one lepton and one neutrino each.
-The $H$ boson decays to $\tau$ pair, each $\tau$ can decay to a lepton a neutrino pair .
-The $H$ boson decays to a $Z$ boson pair, each of which can decay to a lepton pair
This, together with the $t\bar{t}$ decay, can produce 2lss or 3l in the final state as well as b-jets. Again, the cross section of $ttH$ ($\sigma_{t\bar{t}H}$ = 508.5 fb calculated in [6]) is much higher when compared to $t\bar{t}t\bar{t}$, which motivates, as in the case of $t\bar{t}Z$ and $t\bar{t}W$, to include in the analysis regions with a high number of b-jets and a large total transverse momentum (defined in 3.1) to reduce the presence of this type of background.

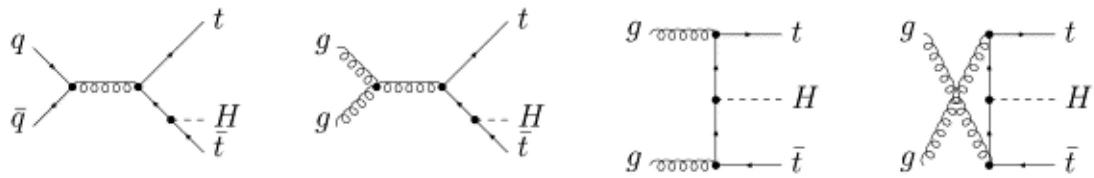


Figure 2.4: A set of Feyman LO diagrams for $t\bar{t}H$ are shown. Image taken from [4]

### 2.2.4 Non-prompt leptons

The two types of regions (2lss and 3l) used during the analysis are based on the number of leptons reconstructed in the final state, these leptons are generally coming from $W$, $Z$ or $\tau$ decay, so it has to be considered whether all the leptons found in the final state are valid (prompt leptons) or come from other sources (non-prompt leptons). For instance, particles produced in the decay of B mesons can give raise to leptons. Therefore, it is important to use a method that allows to select leptons coming from prompt $W$ or $Z$ boson decays efficiently and to minimise the number of non-prompt that are taken into account. The lepton selection used during this analysis is detailed in section 3.2.

### 2.2.5 Other Backgrounds

The main backgrounds expected during the analysis were described above, but this does not mean that they will be the only ones. However, those described in section 2 are the ones that are expected to be encountered in the greatest numbers and for which the event selection criteria described in 3.3.1 and 3.4.1 have been designed as such. Those not mentioned are considered less relevant to the analysis as they will generally appear in smaller quantities. Some examples are: $WZ$, $ZZ$, and photon conversions. There are small contributions from other processes with a low cross section, which we label as “Rares”.

# 3. Event Selection

## 3.1 General Aspects

The objective is to design an analysis to study $t\bar{t}t\bar{t}$ production using pp collisions in the CMS detector. To do that, we use collision simulations that emulate the detector conditions in the 2018 data taking. The simulations will be studied assuming that the $t\bar{t}t\bar{t}$ process really exists as predicted within the SM. Assuming that the data that will be observed is equal to the simulations, we optimize the analysis, trying to obtain the best sensitivity to the process.
For this purpose, two types of regions will be used: 2lss and 3l regions.
The regions 2lss concentrate more signal, and therefore are bring the larger sensitivity to the analysis, however adding regions 3l allows to increase the amount of data to be analysed, enhancing the sensitivity to the process. The regions in the analysis seek a balance between signal and the amount of data analysed. If only regions with a large number of b-jets and total transverse momentum were used, even though a lot of signal compared to the amount of background would be found. However, the absolute number of signal events would be very small and the statistical error would be very high, yielding to a low sensitivity. Therefore, regions with a low number of b-jets and low total transverse momentum are also used to increase the number of events analysed because even though they introduce a larger amount of background, the statistical error is reduced.
To optimize the analysis, regions were constructed using three variables: number of b-jets loose (nbLoose), number of b-jets medium (nbMedium) and total transverse momentum (ht):
-**nbLoose**: Number of b-jets detected in the final state with a 10% mistag rate (probability of identifying a light flavor jet as a b-jet).
-**nbMedium**: Number of b-jets detected in the final state with a mistag rate of 1%.
-**ht**: Scalar sum of the transverse momentum of all final state particles in the event.
Using combinations of these three variables and the event selection criteria described in 3.3.1 and 3.4.1, all regions of the analysis are constructed.

## 3.2 Lepton Selection

One of the bases of the analysis is the presence of two leptons with the same sign in the final state, which is why it is key to correctly identify the leptons. In this case the prompt leptons come from the decay of $W$ bosons, $Z$ bosons in the case of $t\bar{t}Z$ and $\tau$ or $W$ in the case of $t\bar{t}H$. On the other hand the non-prompt leptons generally come from jets, b-jets generally in this case. Therefore, variables must be used to distinguish leptons produced by W,Z or $\tau$ decays from those inside jets.
Since non-prompt leptons are usually found inside jets, they are often accompanied by other objects. Isolation variables are used to distinguish between them,so a cone is constructed around the lepton and the relative particle flow (PF) isola-

tion is computed according to the following expression[11]:

$$I = \sum p_T^i / p_T^{lep} \tag{3.1}$$

Where the sum runs for all particles with a $p_T^i$ inside the previously constructed cone, produced in the main primary vertex and with a $p_T^{lep}$ corresponding to the $p_T$ of the lepton.
The use of isolation variables is key, since otherwise during the analysis many non-prompt leptons would be detected as prompt lepton, because they are really leptons and other variables used for discriminating leptons from other particles are not useful. These non-prompt leptons do not come from the decay of a $Z$,$W$ boson or a $\tau$, making the lepton counting in the final state a big mess, which will invalidate the analysis.
In previous analysis of $t\bar{t}t\bar{t}$ such us the one in [2] the lepton selection used in the analysis is based in the use of isolation variables and impact parameter variables. This analysis also uses impact parameter related variables to distinguish between prompt leptons and non prompt leptons. If the lepton is prompt it will come directly from the primary vertex. However, if it was produced by a particle with a finite half-life, the lepton will come from another point slightly away from the primary vertex.
This thesis uses for the first time in a $t\bar{t}t\bar{t}$ analysis a Boosted Decision Tree (BDT) to further improve prompt lepton detection and therefore the results of the analysis. The BDT trained using data from simulations. It is trained separately for electrons and muons by simulating detector conditions during 2016 and 2017. This thesis uses simulations from 2018 but since the detector conditions didn't change significantly the training is still valid. The variables used by the BDT are of several types: kinematic, isolation, b-tagging, identification and impact parameter variables[11].

## 3.3 2lss Regions Construction

### 3.3.1 Event Selection Criteria for 2lss

In all 2lss regions, in addition to the variables used for defining the regions, an event selection criteria is applied to enhance the analysis, it is detailed below:

- Events are only considered if they contain a pair of leptons with an invariant mass greater than 12 GeV, to avoid low mass resonances
- Events are only considered if there are 2 leptons in the final state
- Events are only considered if the momentum of the two leptons is greater than 15 and 25 GeV respectively, to reduce some background and have a better selection of events than the default trigger one.
- Events are only considered if the 2 leptons are tight (selection criteria from 3.2), to avoid counting non-prompt lepton
- Events are only considered if the 2 leptons have the same charge, to ensure 2lss.

- Events with two electrons are rejected if their invariant mass is consistent with that of the $Z$ peak, to avoid backgrounds that involve $Z$ boson such us $t\bar{t}Z$.
- Events with opposite sign same flavor lepton pair near the $Z$ peak are rejected, to avoid backgrounds such us $t\bar{t}Z$.
- If both leptons are electrons, reject $Ze\bar{e}$ events (because the $e\bar{e}$ pair is consistent with the $Z$ peak) and events with a missing transverse momentum larger than 30 GeV. This allows to reduce $Z \to ee$ events, in which the charge of an electron was not correctly measured.
- An algorithm is used to reject electrons that have been produced in the interaction between a photon and the detector. Additionally, the electron track must not be missing any hits and the different measurements of the electron charge must be compatible.
- Events are only considered if the uncertainty in the muon momentum is less than 20%, to avoid mouns which charge was not correctly measured

Using the variables and event selection criteria previously described, the 2lss regions shown in the table 3.1 were constructed.

## 3.4 3l Regions Construction

### 3.4.1 Event Selection Criteria for 3l

As with regions 2lss, for regions 3l an event selection criteria is applied in order to improve the analysis:

- Events are only considered if they contain a pair of leptons with an invariant mass greater than 12 GeV, to avoid low mass resonances
- Events are only considered if there are 3 leptons in the final state, to ensure 3l
- Events are only considered if the 3 leptons are tight (selection criteria from 3.2), to ensure counting prompt leptons
- Events are only considered if the momentum of the three leptons is greater than 25, 15 and 10 GeV respectively, to reduce some background and have a better selection of events than the default trigger one.
- Events with opposite sign same flavor lepton near $Z$ peak are rejected, to avoid backgrounds such us $t\bar{t}Z$
- An algorithm is used to reject electrons that have been produced in the interaction between a photon and the detector. Additionally, the electron track must not be missing any hits.

Using the variables and the event selection criteria previously described, the 3l regions shown in the table 3.2 were constructed.

| nbLoose | nbMedium | ht (GeV) | Region |
|---|---|---|---|
| 2 | $<2$ | $\leq 300$ | 1 |
| | | $300<ht\leq 400$ | 2 |
| | | $400<ht\leq 500$ | 3 |
| | | $>500$ | 4 |
| | 2 | $\leq 300$ | 5 |
| | | $300<ht\leq 400$ | 6 |
| | | $400<ht\leq 500$ | 7 |
| | | $500<ht\leq 600$ | 8 |
| | | $>600$ | 9 |
| 3 | $<3$ | $\leq 300$ | 10 |
| | | $300<ht\leq 400$ | 11 |
| | | $400<ht\leq 500$ | 12 |
| | | $500<ht\leq 600$ | 13 |
| | | $>600$ | 14 |
| | 3 | $\leq 300$ | 15 |
| | | $300<ht\leq 400$ | 16 |
| | | $400<ht\leq 500$ | 17 |
| | | $500<ht\leq 600$ | 18 |
| | | $>600$ | 19 |
| $\geq 4$ | $<2$ | $\leq 300$ | 20 |
| | | $300<ht\leq 400$ | 21 |
| | | $400<ht\leq 500$ | 22 |
| | | $500<ht\leq 600$ | 23 |
| | | $>600$ | 24 |
| | 2 | $\leq 300$ | 25 |
| | | $300<ht\leq 400$ | 26 |
| | | $400<ht\leq 500$ | 27 |
| | | $500<ht\leq 600$ | 28 |
| | | $>600$ | 29 |
| | 3 | $\leq 300$ | 30 |
| | | $300<ht\leq 400$ | 31 |
| | | $400<ht\leq 500$ | 32 |
| | | $500<ht\leq 600$ | 33 |
| | | $>600$ | 34 |
| | $\geq 4$ | $\leq 300$ | 35 |
| | | $300<ht\leq 400$ | 36 |
| | | $400<ht\leq 500$ | 37 |
| | | $>500$ | 38 |

Table 3.1: The 2lss regions constructed for the analysis are shown.

| nbLoose | nbMedium | ht (GeV) | Region |
|---|---|---|---|
| 2 | <2 | ≤300 | 1 |
| 2 | <2 | 300<ht≤400 | 2 |
| 2 | <2 | 400<ht≤500 | 3 |
| 2 | <2 | >500 | 4 |
| 2 | 2 | ≤300 | 5 |
| 2 | 2 | 300<ht≤400 | 6 |
| 2 | 2 | 400<ht≤500 | 7 |
| 2 | 2 | >500 | 8 |
| 3 | <3 | ≤300 | 9 |
| 3 | <3 | 300<ht≤400 | 10 |
| 3 | <3 | 400<ht≤500 | 11 |
| 3 | <3 | >500 | 12 |
| 3 | 3 | ≤300 | 13 |
| 3 | 3 | 300<ht≤400 | 14 |
| 3 | 3 | 400<ht≤500 | 15 |
| 3 | 3 | >500 | 16 |
| ≥4 | <2 | ≤300 | 17 |
| ≥4 | <2 | 300<ht≤400 | 18 |
| ≥4 | <2 | 400<ht≤500 | 19 |
| ≥4 | <2 | >500 | 20 |
| ≥4 | 2 | ≤300 | 21 |
| ≥4 | 2 | 300<ht≤400 | 22 |
| ≥4 | 2 | 400<ht≤500 | 23 |
| ≥4 | 2 | >500 | 24 |
| ≥4 | 3 | ≤300 | 25 |
| ≥4 | 3 | 300<ht≤400 | 26 |
| ≥4 | 3 | 400<ht≤500 | 27 |
| ≥4 | 3 | >500 | 28 |
| ≥4 | ≥4 | ≤300 | 29 |
| ≥4 | ≥4 | 300<ht≤400 | 30 |
| ≥4 | ≥4 | 400<ht≤500 | 31 |
| ≥4 | ≥4 | >500 | 32 |

Table 3.2: The 3l regions constructed for the analysis are shown

## 3.5 Data processing

Once all the regions have been constructed, each region is analysed individually, obtaining the number of signal and background events in that region.
After analysing all the regions individually, the information obtained from each region is combined. Using the sum of all regions the likelihood function is computed, making a maximum likelihood estimation that best fits the data and then computing the significance of the results.

# 4. Results

## 4.1 Expected Number of Events

Analysing the regions in tables 3.1 and 3.2 individually gives the contribution of each process to that region, these results are combined and shown in figures 4.1 (regions 2lss) and 4.2 (regions 3l).

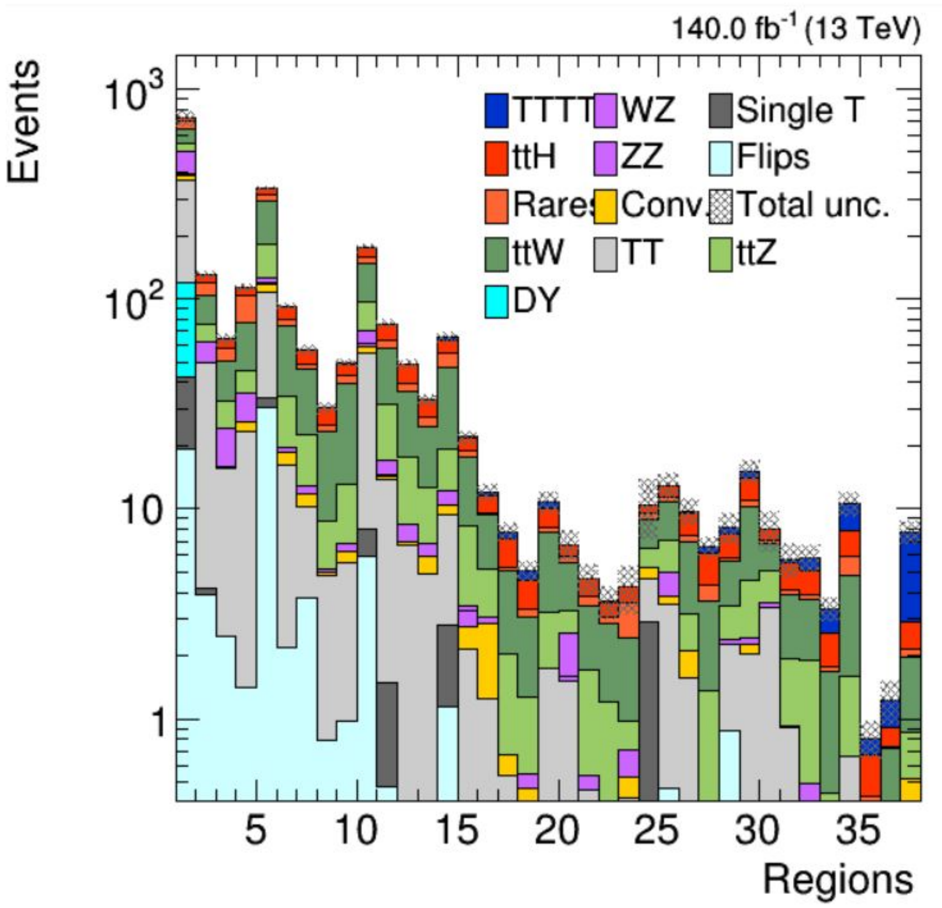


Figure 4.1: The number of events in logarithmic scale of each type found in the 2lss regions from table 3.1 is shown.

Figures 4.1 and 4.2 show that, despite applying the event selection criteria described in sections 3.3 and 3.4, the amount of background obtained is much larger than the amount of signal, which was expected since, as mentioned in section 2, the cross section of the main backgrounds is much bigger than the $t\bar{t}t\bar{t}$ one.
In figure 4.1 it can be see that, when the number of b-jets and total transverse momentum is large as, for example, in region 38 from table 3.1, the background largely decreases as the number of $t\bar{t}t\bar{t}$ events increases. As expected, the number of total events decreases in comparison to low number of b-jets and low total transverse momentum regions such us region 1 from table 3.1, however as its seen in the figure these regions concentrate a lot of background and little signal.

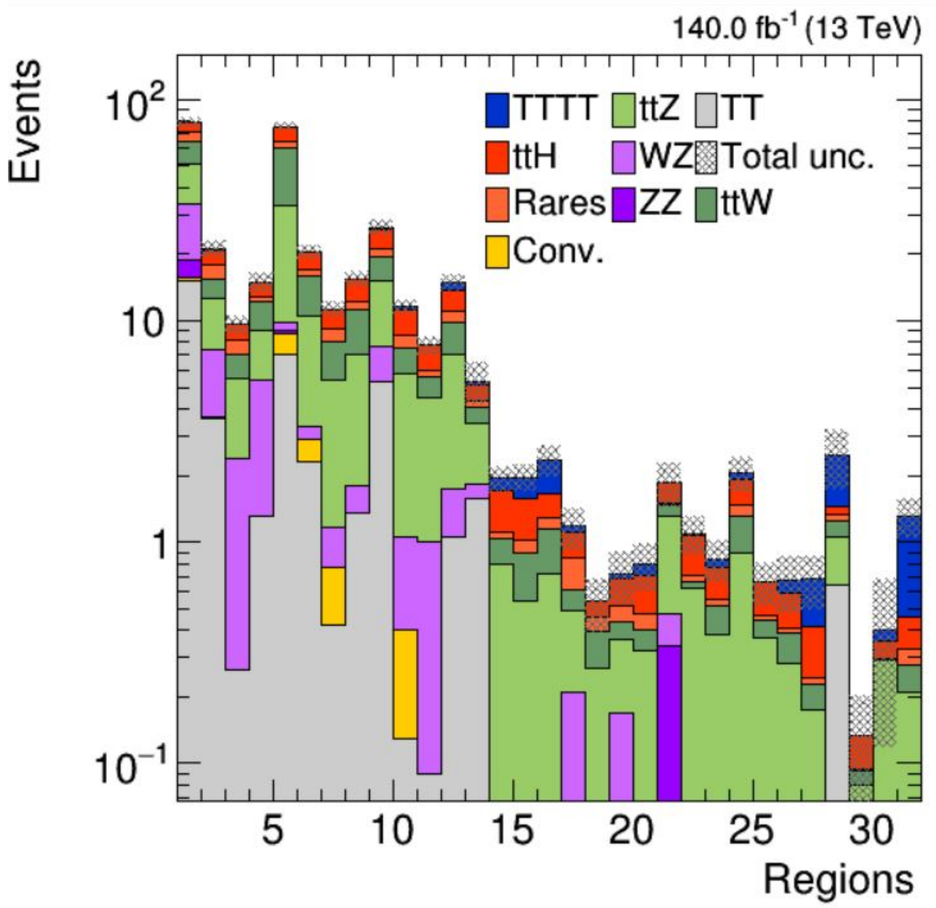


Figure 4.2: The number of events in logarithmic scale of each type found in the 3l regions from table 3.2 is shown.

## 4.2 Measurement of the Signal Strength

In this section, we aim to measure the signal production cross section and compare it with the expectations. To do that, we define the signal strength as:

$$r = \frac{\sigma_{Measured}}{\sigma_{SM}} \tag{4.1}$$

Where $\sigma_{Measured}$ represents the measured cross section and $\sigma_{SM}$, the cross section predicted by the SM.
We use the information from all the event counts shown in figure 4.1 to determine $r$. One way to measure the signal strength of the process being studied is to perform a Maximum Likelihood Estimation which gives a measure of how compatible the results are with the model you are looking for, in this case the model is $t\bar{t}t\bar{t}$ as described in the SM.
The method consists in adjusting the parameters of a function known as the likelihood function in such a way that it maximises the likelihood of the process described by the model having produced the data observed.
Applied to this analysis, the maximum likelihood fit gives the value of the parameter $r$ from the likelihood function. So, the closer to 1, the more compatible the cross section of the measured process is with the prediction.
The adjustment of the parameter $r$, using 2018 Monte Carlo simulations data ,

gives the following results:

$$r = 1.00^{+0.41}_{-0.48} \tag{4.2}$$

It was expected to obtain r=1 since, as stated in section 3.1, in this analysis we have been assumed that the process exists and simulations that generate the process according to the SM have been analysed. It is interesting to see with what uncertainty the signal strength has been obtained, .This total uncertainty in the fit comes from the statistical uncertainty of the observed data and from various systematic error sources, for example the uncertainty in the lepton detection efficiency or the uncertainty in the trigger efficiency. As will be seen in the next section, not all of them contribute equally and the uncertainty associated with b-tagging and the uncertainty associated with the MC statistics stand out.

## 4.3 Analysis Sensitivity

One way to test the sensitivity of the analysis to $t\bar{t}t\bar{t}$ is to calculate the p-value. The p-value is the probability of obtaining $r = 1$ assuming $r = 0$ (the process under study does not exist). So the smaller the p-value, the more sensible is the analysis to $t\bar{t}t\bar{t}$ . However, it can be a bit inconvenient to be working with small p-values, so the significance is calculated instead.
Significance (also named z-value) is related to the p-value as follows:

$$P(x > z - value) = p - value \tag{4.3}$$

Where $P(x)$ is the probability in a Gaussian distribution. This way we work with standard deviations instead of p-values, which is more intuitive.
Combining the results for regions 2lss and 3l shown in figures 4.1 and 4.2, the significance shown in the table 4.1 is obtained.
To put this result into perspective, it should be noted that a significance of 2.8 standard deviations is equivalent to a p-value of $p \approx 0.0026$, which means the probability of measuring $r = 1$ when the true value is $r = 0$ ($t\bar{t}t\bar{t}$ does not exist). This way, we can ensure that the $t\bar{t}t\bar{t}$ events shown in figures 4.1 and 4.2, are indeed $t\bar{t}t\bar{t}$ and not background with a 2.8 standard deviations confidence level (It can never be fully guaranteed).
It has to be taken into account that the significance of the results is affected by the different sources of uncertainty (uncertainties in the efficiency of lepton detection, uncertainties associated with b-tagging...).
If the different sources of uncertainty are removed from the significance calculation, it can be seen how limited the analysis is by the sources of uncertainty and what possible improvements could be applied in the future to improve it. By recalculating the significance while removing the different uncertainties it is obtained the result shown in table 4.1
Considering that 3.4 standard deviations equates to a p-value of $p \approx 0.0003$, the analysis still has great room for improvement only by reducing the systematic uncertainties.
In order to reduce uncertainties in future analyses, it must be taken into account that there are a large number of sources of uncertainty and not all of them affect the significance of the results in the same way.

For this reason, it is interesting to recalculate the significance excluding the different sources of uncertainty one by one, so that it can be seen which ones affect the results the most. The results are shown in table 4.1. After doing this, two main sources of uncertainty are obtained: the uncertainties associated with the b-tagging and the uncertainties associated with the Monte Carlo statistics used in the simulation.
Moreover, the analysis is based on 2lss regions and 3l regions, one way to check that it is really optimal to use both types and that it would not produce better results to stay with only one of the two is to recalculate the significance of the results separately for 2lss and 3l. The results are shown in table 4.1. It can be seen that most of the sensitivity of the analysis is due to 2lss regions. However, adding 3l regions helps to increase the overall sensitivity of the analysis.

| Calculation | Expected sensitivity (standard deviations) |
|---|---|
| All uncertainties | 2.8 |
| No Uncertainties | 3.4 |
| 2lss | 2.3 |
| 3l | 1.7 |
| B-tagging | 3.0 |
| MC statistics | 3.0 |
| Lepton efficiency | 2.8 |
| Jet resolution | 2.8 |
| Trigger efficiency | 2.8 |
| Dibosons & Conversions Modeling | 2.8 |
| B-tagging efficiency | 2.8 |
| Pile up | 2.8 |
| Kinematic distributions of processes estimated with MC simulations | 2.8 |
| ttH cross section | 2.8 |
| ttW cross section | 2.8 |
| ttZ cross section | 2.8 |
| tHW, tHq and ttWW cross section | 2.8 |
| Luminosity | 2.8 |
| Protons pdf | 2.8 |

Table 4.1: The different calculated significances are shown.

# Conclusions

The aim of the analysis was to obtain the highest sensitivity to the $t\bar{t}t\bar{t}$ process by studying the simulations as if they were experimental data. A significance of 2.8 standard deviations was obtained.
For the analysis, regions with two leptons of the same sign and three leptons in the final state have been used. As has been seen when calculating the significance of the two types of regions separately, both contribute to the final result, although the region with two leptons of the same sign has more significance as it has a larger amount of signal.
Of all the sources of uncertainty present in the analysis, it has been observed that the following stand out: the uncertainties associated with b-tagging and the uncertainties associated with MC statistics.
For future work, there are different ways in which this analysis can be improved. One of them is to reduce uncertainties. The uncertainty associated with the CM statistics could be reduced by increasing the number of events analysed by running more simulations or by adding simulations from other years.
Another way to improve the analysis is to improve the lepton selection, since the analysis is based on the detection of leptons in the final state, improving the precision with which prompt leptons and non-prompt leptons are distinguished may lead to better results.
Optimising more variables when constructing the regions can also contribute to better results, e.g. the number of jets can be added.
The final goal of this analysis is to perform this analysis using experimental data, and performing a real search for $t\bar{t}t\bar{t}$ production.

# Bibliography


[1] M. et al Aaboud. “Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$TeV with the ATLAS detector”. In: *Phys. Rev. D* 99 (7 Apr. 2019), p. 072009. DOI: 10.1103/PhysRevD.99.072009. URL: https://link.aps.org/doi/10.1103/PhysRevD.99.072009.

[2] Morad Aaboud et al. “Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in p p collisions at $\sqrt{s}$= 13 TeV with the ATLAS detector”. In: *Physical review D* 99.5 (2019), p. 052009.

[3] A. M. Sirunyan et al. “Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$TeV”. In: *Journal of High Energy Physics* 2018.8 (Aug. 2018). DOI: 10.1007/jhep08(2018)011. URL: https://doi.org/10.1007%2Fjhep08%282018%29011.

[4] W. Beenakker et al. “NLO QCD corrections to $t\bar{t}H$ production in hadron collisions”. In: *Nuclear Physics B* 653.1 (2003), pp. 151–203. ISSN: 0550-3213. DOI: https://doi.org/10.1016/S0550-3213(03)00044-0. URL: https://www.sciencedirect.com/science/article/pii/S0550321303000440.

[5] A.B. Arbuzov. “Quantum Field Theory and the Electroweak Standard Model”. In: (Jan. 2018). 35 pages, 1–34. 35 p. DOI: 10.23730/CYRSP-2017-004.1. arXiv: 1801.05670. URL: https://cds.cern.ch/record/2315477.

[6] *CERN Yellow Report Page at $\sqrt{s} = 13-14$ TeV*. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV2014#SM_Higgs_production_cross_sectio. Accessed: 2022-08-27.

[7] *Detector CMS experiment*. https://cms.cern/detector. Accessed: 2022-06-11.

[8] Brian Martin. “Nuclear and Particle Physics An Introduction Second Edition”. In: (2021).

[9] Aaron Paul O’Neill et al. “Measurement of the the tttt production cross section in pp collisions at $\sqrt{s}$= 13 TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2021.11 (2021), p. 118.

[10] Sirunyan. “Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV.” In: *Eur. Phys. J. C* 80 (Aug. 2019), 75. 31 p. DOI: 10.1140/epjc/s10052-019-7593-7. arXiv: 1908.06463. URL: https://cds.cern.ch/record/2686802.

[11] Albert M Sirunyan et al. “Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s}$= 13 TeV”. In: *The European Physical Journal C* 81.4 (2021), pp. 1–51.

[12] *Taking a closer look at LHC*. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_layout/idioma/en_GB. Accessed: 2022-06-11.