LIP-STUDENTS-22-06

Study of central and exclusive production of tau-tau pairs at LHC

Alexandre André[1,a] and Pedro Batista[1,b]

[1] *Faculdade de Ciências da Universidade de Lisboa, Portugal*

Project supervisor: Matteo Pisano, Michele Gallinaro

October 17, 2022

Abstract. A study of the central production of tau-tau pairs in proton-proton collisions. The study was performed using simulated samples for the Ttjets and Drell-Yan backgrounds and for the signal events. We did a careful analysis of the QCD background using a data driven approach and confirmed that our method of estimation was valid for this background. Control Regions revealed that our simulated samples were in good accordance with the expectations, meaning that the simulations and the approximation for the QCD were both well made. We used a Multivariate Analysis Tool (TMVA) to separate the backgrounds from the signal and analyzed the resulting plots. In the end, we were able to derive a limit on the cross section for the signal event.

Keywords: LHC, proton-proton collision, tau-tau exclusive, exclusive production, TMVA

1 Introduction

1.1 The CMS detector

The "Large Hadron Collider" (LHC) [1] is a particle accelerator near Geneva and it allows to accelerate protons to an energy of 6.5 TeV and a velocity close to the speed of light. In this study, the protons collide in interaction point 5 (IP5) of LHC. Around IP5 there is a particle detector called "Compact Muon Solenoid" (CMS) [2], figure 1. The CMS allows the measurement of kinematic quantities (momentum, energy, ...) of the particles produced in the interaction. A huge solenoid magnet is integrated in the CMS. This takes the form of a cylindrical coil of superconducting cable that generates a field of 4 tesla. The field is confined by a steel "yoke" that forms the bulk of the detector's 14000-tonne weight. At a distance of ±200m from IP5, on each side of CMS, is located a set of detectors called "Precision Proton Spectrometer" (PPS) [3], presented in figure 2, that are capable of detecting the protons resulting from the collision and are able to measure the angle by which they deviate from the main axis of the collision. The data collected in the PPS about the final product protons is stored separately from the data collected in the CMS. This data will be used in a data processing step called "proton enrichment", explained later.

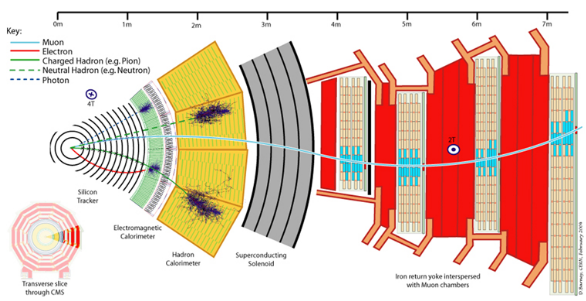

Figure 1. Schematic transverse view of the CMS detector

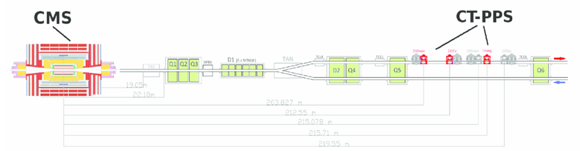

Figure 2. Side view of the PPS detector

1.2 Relevant variables

In this subsection we present a list of variables that can be measured by the detector and are used for the analysis in this work.

- θ : polar angle of the trajectory of a particle with respect to the counterclockwise proton beam;
- $\eta = -\ln(\tan\theta/2)$: pseudorapidity - is in simple terms the angle of a particle relative to the beam axis;
- P_t : transverse momentum;
- M_t : invariant transverse mass;
- $aco = \frac{|\Delta\phi|}{\pi}$, where $\Delta\phi = \phi_2 - \phi_1$: acoplanarity - "angle" between the two particles in the final state and is a number in the range of 0 to 1;
- $bjets$: the number of bottom quarks originated from the decay of the top quark;
- $\sqrt{S}$: initial energy of the protons, $13000GeV$;
- $\xi = \frac{|P_{t,i}|-|P_{t,f}|}{|P_{t,i}|}$: momentum lost by the protons - measured by the PPS;

1.3 Proton-proton collisions

Protons are not elementary particles [4], therefore, during a collision, they may dissociate. During the interactions, just a pair of quarks or gluons interact, this is called primary interaction, while the remaining objects originate "jets" in the final state, as shown in figure 3. We will call "signal event" the process we want to study. In signal events the two incoming protons don't dissociate, instead they lose energy and create a tau and anti-tau lepton pair,

[a] e-mail: alexandre_andre@outlook.pt

[b] e-mail: cbatista.pedro@gmail.com

as shown in figure 4. A process characterised by undissociated protons is called "exclusive process". An introduction to such processes, including a detailed state of the art, can be found in Ref. [5]. In our case of study one tau decays into a muon (μ), plus a tau-neutrino and a muon-antineutrino; or into an electron, plus a tau-neutrino and an electron-antineutrino, and the other tau decays hadronically (τ_h). Then there are three main background processes that are competitive, those being Ttjets, Drell–Yan and QCD. Ttjets is when the interaction between the 2 quarks result in a pair of top and anti-top quarks that decay into a bottom quark and other particles. The Drell-Yan background has two possibilities because the interaction can result in a pair of lepton and anti-lepton or it can result in a Z boson that then decays to a pair of lepton and anti-lepton. The QCD background is more complex and hard to explain but what happens is that the interaction generates hadron jets and it doesn't generate well defined particles.

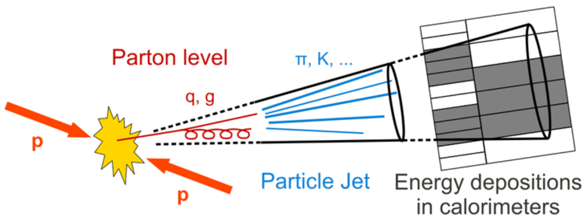

Figure 3. Formation of jets of particles.

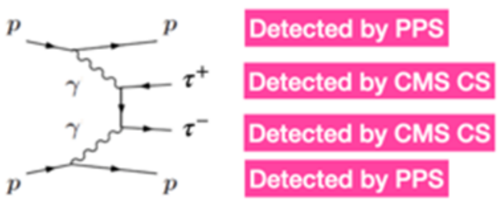

Figure 4. Feynman diagram for the signal event.

2 Sample processing

Our analysis is based on the study of a luminosity of 40 fb^{-1}; we studied the case in which one tau decays into a muon and the other one hadronically. Our samples were subject to a Trigger in order to guarantee that our signal event occurred and to only save the relevant information about the event. The Trigger that was used guaranteed the presence of an isolated muon with transverse momentum (P_t) higher than 24 GeV.

The set of samples used is composed of a sample of data and simulated Monte Carlo. The simulated samples are the signal event and the Ttjets and Drell-Yan backgrounds. We have to work with simulated samples because it is very hard to separate this events directly from the sample of data. Because of this we must calculate the weight of the simulated samples. Then we filtered these samples to verify that the kinematic properties were respected. This is related with the fact that the PPS can only measure protons that result from interactions with an invariant mass of 300 GeV or higher. On top of that we ask for the transverse momentum of the tau to be at least 100 GeV and at least 35 GeV for the muon. We also had to do a skimming related to the geometric acceptance because if the value of the pseudorapidity (η) was greater than 2.4 then the particles wouldn't be detected by the CMS. The other type of skimming that we did was to make sure that the electrical charge of the particles in the final state were opposite to each other. We did this by conditioning the events so that the product of the sign of each particle in the final state was negative.

The QCD background is the only one studied by looking at a sample of data, this is due to the fact that this background is very hardly defined and because of this it's not possible to simulate it with great precision. To resolve this problem we do a skimming of a sample of data in which we only choose the events where the particles in the final state have the same sign. This is done because we know that both the Ttjets and Drell-Yan backgrounds, as well as the signal event, generate particles in the final state with opposite charges so, by doing this skimming, we isolate the events that are part of the QCD background. In theory, since the QCD background appears to be random and generate non defined jets of particles, there must be approximately 50% of particles in the end state with opposite charges and another 50% with the same charge present in the real set of data and similarly there must be as much background with the same charge as background with opposite charges. With this in mind, we get a very good approximation of the QCD background. We will show that this approximation is very good in a later analysis. Table 5 shows the amount of events that we had after each cut, normalized to 40 fb^{-1}.

	Signal	QCD	DY	ttjets
#events tau and muon	9.4	2.7 x 10^5	7.7 x 10^4	1.3 x 10^4
#events tau and muon (opposite charge)	9.2	5.7 x 10^4	7.4 x 10^4	1.2 x 10^4
#events Pt(tau)>100Gev; Pt(muon)>35Gev	4.4	2.6 x 10^3	1.3 x 10^3	2.0 x 10^3
|η|<2.4	4.4	2.4 x 10^3	1.3 x 10^3	2.0 x 10^3

Figure 5. Table of events after each kinematic cut, normalized to the luminosity of 40fb^{-1}.

3 Kinematic distributions

We decided to plot some kinematic distributions of background and signal samples. These distributions were stored in three histograms. These histograms were used to determine the best cuts to be made for the control regions.

3.1 Histograms of variables

In the first histogram (figure 6) we have acoplanarity. Acoplanarity represents the angle between the 2 particles divided by π and is a number in the range of 0 to 1. There is also 1 special thing to notice which are the 2 peaks of the Drell-Yan background. These are due to its 2 possible decays, the higher peak close to acoplanarity 1 is when the protons collide and create the pair of lepton and anti-lepton which go in opposite directions, the lower peak is from the process when the protons originate a Z boson which has a direction of production and the leptons that result from the decay of the Z boson will follow the same general direction making the angle between them smaller and therefore having lower acoplanarity.

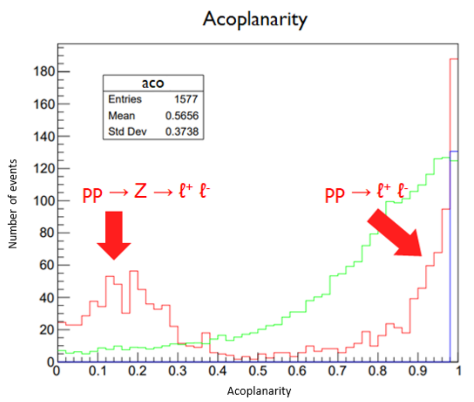

Figure 6. Histogram of acoplanarity for the Ttjets background (green), Drell-Yan background (red) and for the signal event (blue). The acoplanarity is calculated based on the ϕ angle of the muon and of the hadronic tau.

The histogram in figure 7 shows the invariant mass of the particles produced and we can see the invariant mass of the top quarks in the Ttjets background, around 200 GeV, and we can also see the invariant mass of the leptons created by the Drell-Yan process around the same value, but specially we can see the invariant mass of the Z boson, close to 90 GeV. It should be noted that in the histogram, the peak for the Z boson is actually a little lower than 90 GeV, this is due to the reconstruction problems of CMS that cant detect all the energy of a particle, therefore, some is lost.

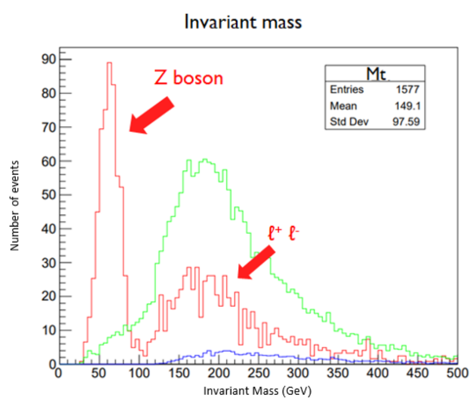

Figure 7. Histogram of the invariant mass of the $\mu\tau_h$ system for the Ttjets background (green), for the Drell-Yan background (red) and for the signal event (blue).

In the last histogram (figure 8) we have the momentum of the particles, and again the Drell-Yan shows 2 peaks, the left one is from the 2 leptons and the right one is from the Z boson.

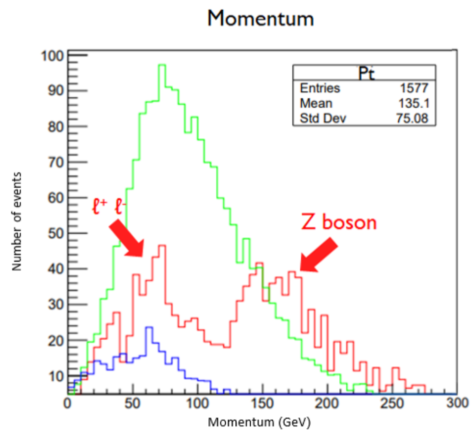

Figure 8. Histogram of the transverse momentum of the $\mu\tau_h$ system for the Ttjets background (green), for the Drell-Yan background (red) and for the signal event (blue).

3.2 Control regions

To verify that the simulated samples are in accordance to the expectations we created several control regions where we tested if each background matched a sub-sample of data. To do this we made some cuts to isolate each background, these were determined using the histograms shown in the previous section. After the cuts were made, we overlapped the backgrounds to see if they matched the expectations.

To isolate the Ttjets background we used the conditions of $bjets \geq 1$ and $aco \geq 0.35$. We obtained control regions with about 60% of Ttjets background (figure 9).

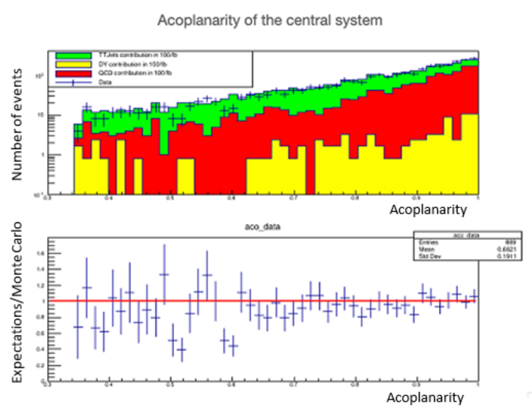

Figure 9. Acoplanarity distribution ($\mu\tau_h$ system) of the Ttjets control region where Ttjets events (green) are $\approx$ 60% of all events, overlapped with Drell-Yan (yellow) and QCD (red). Below is shown the ratio between the expectations and sum of all backgrounds.

For Drell-Yan we used the conditions $M_t \leq 100GeV$ and $aco \leq 0.35$. Control regions were obtained with about 80% of Drell-Yan background (figure 10).

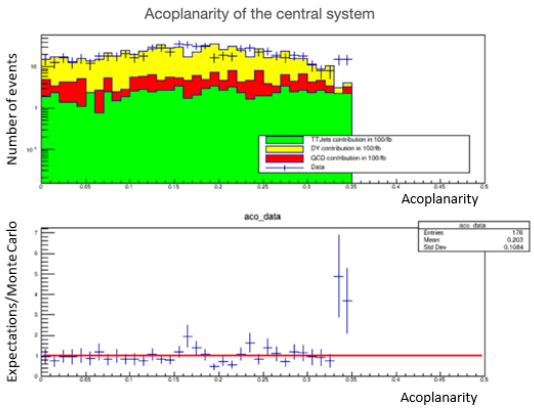

Figure 10. Acoplanarity distribution ($\mu\tau_h$ system) of the Drell-Yan control region where Drell-Yan events (yellow) are $\approx$ 80% of all events, overlapped with Ttjets (green) and QCD (red). Below is shown the ratio between the expectations and sum of all backgrounds.

For QCD we used the conditions $M_t \geq 300GeV$ and $bjets < 1$. Control regions were obtained with about 90% of QCD background (figure 11).

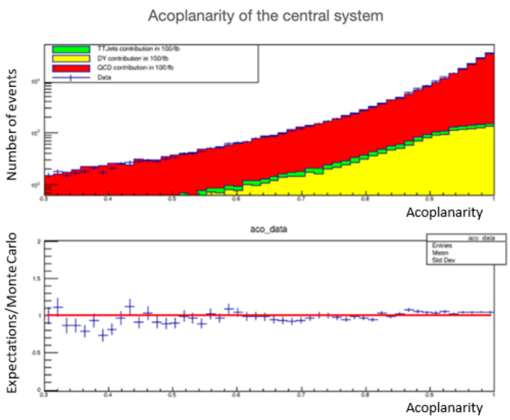

Figure 11. Acoplanarity distribution ($\mu\tau_h$ system) of the QCD control region where QCD (red) are $\approx$ 90% of all events, overlapped with Ttjets (green) and Drell-Yan (yellow). Below is shown the ratio between the expectations and sum of all backgrounds.

If we analyze these control regions we can see that there is a good agreement between expectations and the simulated samples, because all the values come very close to the horizontal line of 1, that represents the theoretical perfect match of the data. This means that the samples we used were simulated correctly and also shows that the approximation made in section 2 was very good.

4 Proton enrichment

After the verification of the Monte Carlo (MC) samples we have to do a last step before we can analyze the results and draw conclusions from them. Simulated samples do not contain pile up protons, therefore, we must enrich them. Since, for background events, protons come from simultaneous and uncorrelated events, it is enough "attaching" to each event a pair of random protons. The enrichment is done according to the probability of having pile up protons. These protons are real protons measured by the PPS. On the contrary, for signal events, there is a very important relation between the invariant mass of the central system and the invariant mass of these protons which is: $M_t = \sqrt{S\xi_1\xi_2}$. For this reason, signal samples contain the ξ information (which has been simulated *ad hoc*).

5 Multivariate analysis

Despite the sample selection mentioned in section 2, the separation between signal and background was not satisfactory: there was more signal than background. To further separate the background from the signal we used a Multivariate Analysis technique. The Multivariate Analysis Tool (TMVA) [6] considers various signal and background distributions, whose variables are discriminant, that is, very different between signal and background. For example, for the Drell-Yan background, the acoplanarity distribution has a peak of 1 for the signal and for the background is a distribution that goes continuously from zero to 1 (figure 14). Another example is the invariant mass matching that is centered around zero for the signal and has a random distribuition for the background (figure 14).

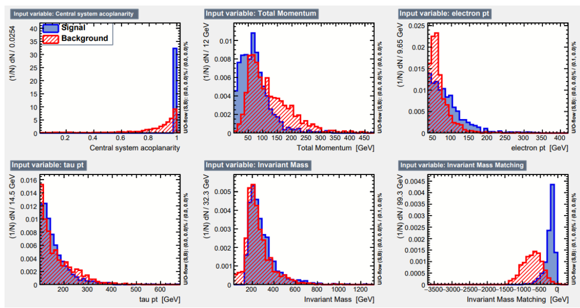

Figure 12. Distribution histograms for some studied observables in the QCD background.

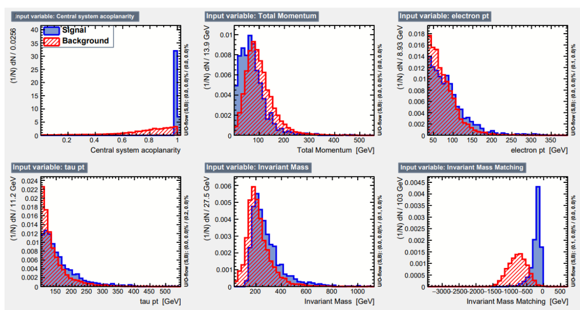

Figure 13. Distribution histograms for some studied observables in the Ttjets background.

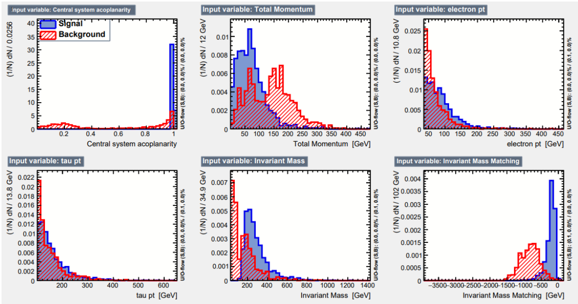

Figure 14. Distribution histograms for some studied observables in the Drell-Yan background.

After considering all the discriminant distributions, the TMVA creates a new one (called test statistic) for each background, which is more discriminating than the input distributions. These distributions can be seen in figure 15.

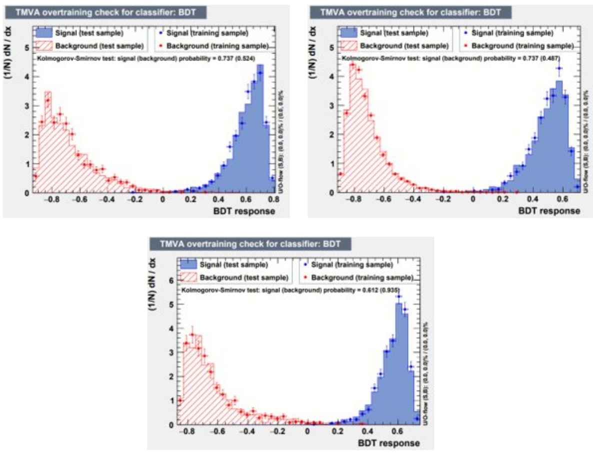

Figure 15. Test statistic distributions for QCD (top left), Ttjets (top right) and Drell-Yan (bottom).

As shown in figure 15, there is an excellent discrimination between the background distribution (that is confined to the left) and the distribution of the signal (that is confined to the right). Nevertheless, this technique is not 100% accurate because there are still some events on the left that are signal and others on the right that are background.

In our case, the TMVA used was Boosted Decision Trees (BDT). In order to choose the best TMVA a graphic analysis was performed to compare two methods available: BDT and MLP (Multilayer Perceptron). For each background we did a plot in which the horizontal axis represents the signal efficiency (i.e. the probability of keeping the signal) and in the vertical axis we have the probability of eliminating the background. The ideal curve is such that the area underneath it is maximised. It happens that such curve is the straigh line $y = 1$. In that case, the probability of eliminating the background is maximal and independent of the signal efficiency. As one can see in figure 16 the BDT was the best method for all backgrounds.

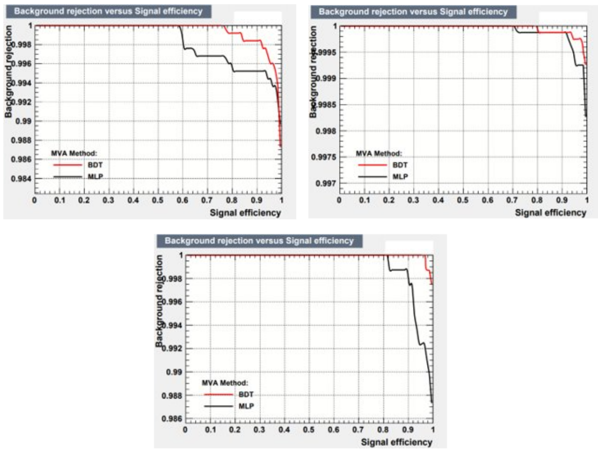

Figure 16. Background rejection versus signal efficiency graphs for the QCD (top left), Ttjets (top right) and Drell-Yan (bottom).

6 Cross section

To achieve the final goal of deriving a limit on the cross section of the MuTau channel, we made use of the Higgs Combine tool. This tool combines the 3 different background distributions from the test statistic (figure 15) and, in doing so, permits to calculate the signal strength, r. It was estimated an upper limit two orders of magnitude higher than the theoretical expectation.

7 Conclusions

In this project we successfully derive a limit on the cross section of the MuTau channel, based on a $40 fb^{-1}$ luminosity sample. In order to do so, we started by simulating samples for the signal event and for the Drell-Yan and Ttjets backgrounds. Then we estimated the QCD background based on a sample of data. The obtained results from the Control Regions showed that the simulations and the QCD estimation were in agreement with the expectations. Following that, we used the TMVA to further separate the backgrounds from the signal, creating the test statistic distributions, which were very discriminant. Finally, we estimated an upper limit for the cross section of this process. This number is about two orders of magnitude higher than the theoretical cross section.

Acknowledgements

We would like to thank our supervisor Matteo Pisano for the unconditional support and the way he received and taught us in such a warm and passionate way. We would also like to thank LIP and the internship organizers who made this wonderful learning experience possible.

References

[1] C. O'Luanaigh, *The Large Hadron Collider*, JINST 3 S08004 (Jan 2014)

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

[3] CMS, TOTEM, *CMS-TOTEM Precision Proton Spectrometer Technical Design Report*, CERN-LHCC-2014-021 (Sep. 2014)

[4] A. De Angelis, M. Pimenta, *Introduction to Particle and Astroparticle Physics*, ISNB 978-8847026872 (2015)

[5] M. Pisano, *Study of central exclusive production of top quark-antiquark pairs at LHC*, DOI: 10.1393/ncc/i2021-21066-9 (2021)

[6] A. Hoecker, *Toolkit for Multivariate Data Analysis with ROOT, Users Guide*, CERN-OPEN-2007-007 (Mar. 2017)