CERN Summer Student Program

Project Report

Feasibility Study for Observation of the $t\bar{t}Z(Z \rightarrow \tau^+\tau^-)$ Process

Ghaida Aldhahri

Supervisors:

Baptiste Ravina

Tomas Dado

August 1, 2025

Contents

1 Introduction

The goal of this project was to investigate the feasibility of measuring the $t\bar{t}Z$ process where the Z boson decays to $\tau^+\tau^-$, a rare and previously unmeasured channel. Our analysis was performed using simulated data corresponding to the full ATLAS Run 2 dataset, representing an integrated luminosity of 140 fb^{-1} at $\sqrt{s} = 13$ TeV. This required distinguishing the signal from dominant background processes such as $t\bar{t}$, $t\bar{t}Z$ (non-$\tau\bar{\tau}$), $t\bar{t}H$, Z+jets, and W+jets.

1.1 Motivation

The top quark pair production in association with a Z boson ($t\bar{t}Z$) provides a fertile ground to study Standard Model processes and rare decays. In particular, we aim to investigate the $t\bar{t}Z$ process where the Z boson decays into a pair of tau leptons ($Z \rightarrow \tau^+\tau^-$), as illustrated in Figure 1.

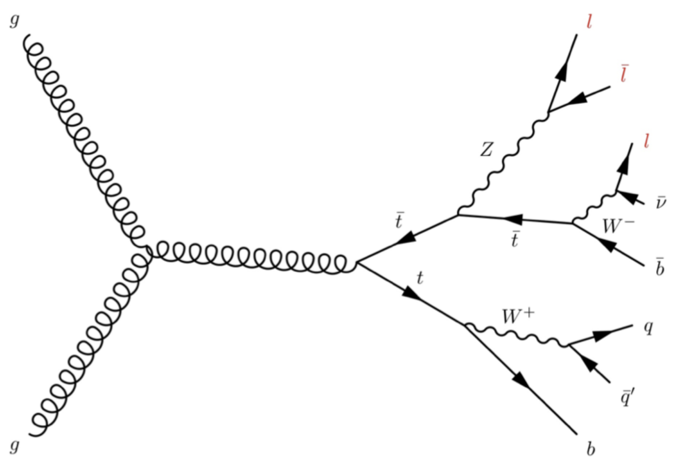

Figure 1: Feynman diagram for $t\bar{t}Z$ production with ($Z \rightarrow \tau^+\tau^-$) and ($t \rightarrow W + b$). This process leads to multi-lepton and multi-jet final states.

While $t\bar{t}Z$ has already been measured in $Z \rightarrow e^+e^-$ or $Z \rightarrow \mu^+\mu^-$ channels, this is the first ATLAS study targeting the $\tau^+\tau^-$ final state (1). This channel is especially compelling because it connects the top quark and the tau lepton — the heaviest quark and heaviest lepton in the Standard Model. Although they do not couple directly, studying this process can provide indirect insight into their interactions. Such measurements may constrain tau lepton behavior, shed light on the structure of mass generation, and potentially reveal signs of physics beyond the Standard Model.

Despite its rarity, this decay channel is important for precision measurements. However, the signal is embedded in complex final states with dominant backgrounds, particularly from $t\bar{t}$ and $t\bar{t}H$ production. Isolating the signal therefore requires careful region selection and background suppression strategies.

The dominant backgrounds considered in this study include $t\bar{t}$, $t\bar{t}H$, Z+jets, and W+jets. These processes are chosen because they can mimic the $t\bar{t}(Z \to \tau^+\tau^-)$ signal through similar final-state signatures: multi-lepton events, jets, and missing transverse energy due to neutrinos. In particular, $t\bar{t}$ is expected to dominate due to its large cross-section and production rate at the LHC, often producing final states with multiple jets and leptons through leptonic W decays. Z+jets and W+jets also produce events with light leptons, taus, and jets, but without producing any top quark and therefore without a peak at the top quark mass in the distribution of the mass of the relevant decay products. $t\bar{t}H$ is structurally similar to $t\bar{t}Z$, but includes a Higgs boson that can decay to leptons or taus and is characterized by a higher invariant mass than the Z boson.

We rely heavily on Monte Carlo simulation to predict how signal and background processes manifest in the ATLAS detector (Figure 2), making it critical to understand the structure of these events and their mimicry by background processes. This motivates our multi-step approach: motivated by Feynman diagrams, we systematically scan through region definitions, analyze the signal and background contributions, apply selections, and compute statistical significance to optimize the analysis.

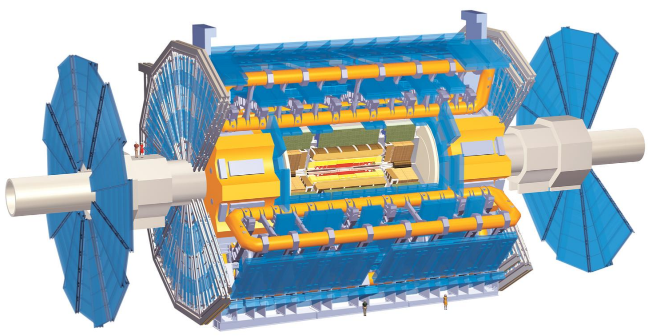

Figure 2: Schematic of ATLAS Detector. © CERN.

2 Methodology

2.1 Defining Signal-like Regions

To isolate the signal, we constructed a set of signal-like regions based on expected decay topologies involving combinations of tau leptons, light leptons, and jets. In our analysis, ”tau” specifically refers to hadronically decaying tau leptons, which are distinguishable from electrons and muons. Since tau leptons can also decay leptonically into electrons or muons, some final states may include multiple light leptons, motivating the inclusion of multi-lepton categories. A summary of the region configurations motivated by the $t\bar{t}Z$ process is shown in Table 1.

	2 tau-jets	1 tau-jet + 1L	0 tau-jets + 2L
0L + 6 jets	2 τ, 6 jets, 0L ✓	1 τ, 6 jets, 1L	0 τ, 6 jets, 2L
1L + 4 jets	2 τ, 4 jets, 1L ✓	1 τ, 4 jets, 2L	0 τ, 4 jets, 3L
2L + 2 jets	2 τ, 2 jets, 2L ✓	1 τ, 2 jets, 3L	0 τ, 2 jets, 4L

Table 1: Sketch of region selection motivated by Feynman diagrams. Rows and columns represent expected final-state multiplicities from $t\bar{t}$ and Z decays. Check marks indicate promising regions that were selected for further analysis.

2.2 Background Composition Across Regions

After defining our signal-like regions, we examined the background composition to assess whether our selections effectively suppress dominant backgrounds while preserving the $t\bar{t}Z$ signal. Background processes were selected because they mimic the topology of the $t\bar{t}Z$ final state. As illustrated in Figure 1, the signal includes top quark decays and a $Z \to \tau^+\tau^-$ decay, which closely resembles the structure of $t\bar{t}$ and $t\bar{t}H$ events, the latter differing only by replacing the Z boson with a Higgs boson. The resulting background composition, visualized using pie charts, confirms these expectations (Figure 3). The dominant background across most regions is $t\bar{t}$, reflecting its large production cross-section at the LHC. Subleading contributions come from $t\bar{t}H$ and Z+jets. The signal fraction is consistently small, underscoring the need for optimized selection cuts to improve signal-to-background separation.

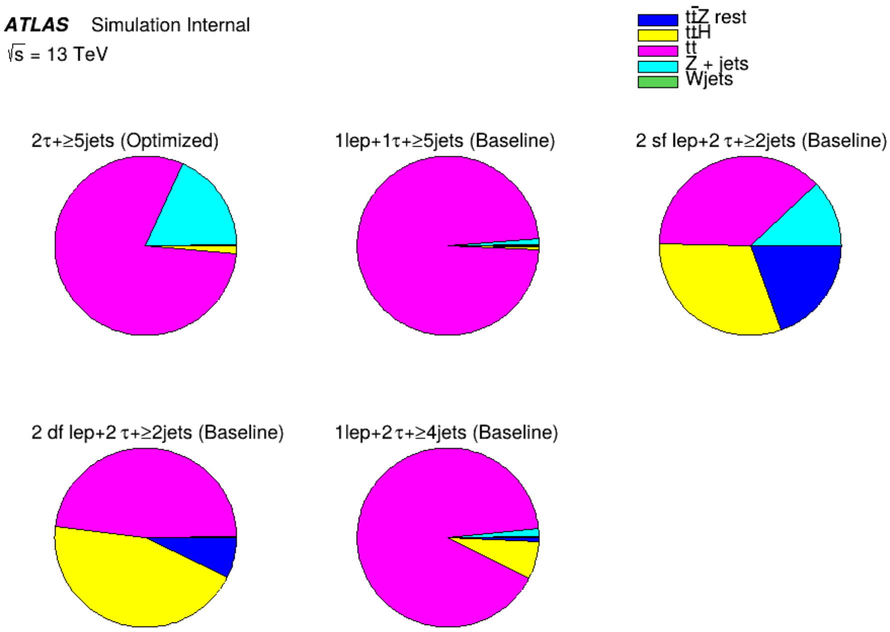

Figure 3: Pie charts showing background composition in each signal-like region. The dominant background across most regions is $t\bar{t}$, consistent with expectations. Signal fraction $t\bar{t}(Z \to \tau^+\tau^-)$ is generally small and must be enhanced via cuts.

2.3 Variable Exploration and Cut Optimization

We initially investigated standard kinematic variables such as jet multiplicity, tau transverse momenta, and missing transverse energy (MET) significance, but these showed limited power in separating signal from background. We also explored mass-related variables, including Missing Mass Calculator (MMC) estimates of the reconstructed Z boson mass (expected near 91 GeV), as well as top quark mass estimates derived from the HyPER machine-learning-based top quark reconstruction algorithm (2; 3).

To improve signal-to-background discrimination, we constructed new derived variables using the MMC algorithm: `"maxw"`, `"mlnu3p"`, and `"mlm"`. These variables represent the invariant mass of a candidate $Z \to \tau^+\tau^-$ decay according to various estimation strategies from the MMC algorithm, under the assumption that all the missing transverse energy in the event originates from the neutrinos associated with the decays of the two τ-leptons. Among these, `"mlm"` exhibited low separation and was excluded from further analysis. Using the more promising `"maxw"` and `"mlnu3p"`, we derived new variables: their geometric mean, arithmetic mean (`amean`), and their difference. Of these, the arithmetic mean proved most effective. We applied selection cuts to this variable in the 0 lepton, 2 taus, and 5 jets region, which showed the highest promise. An example is shown in Figure 4, where the optimized selection based on `amean` significantly reduces background while retaining most of the signal.

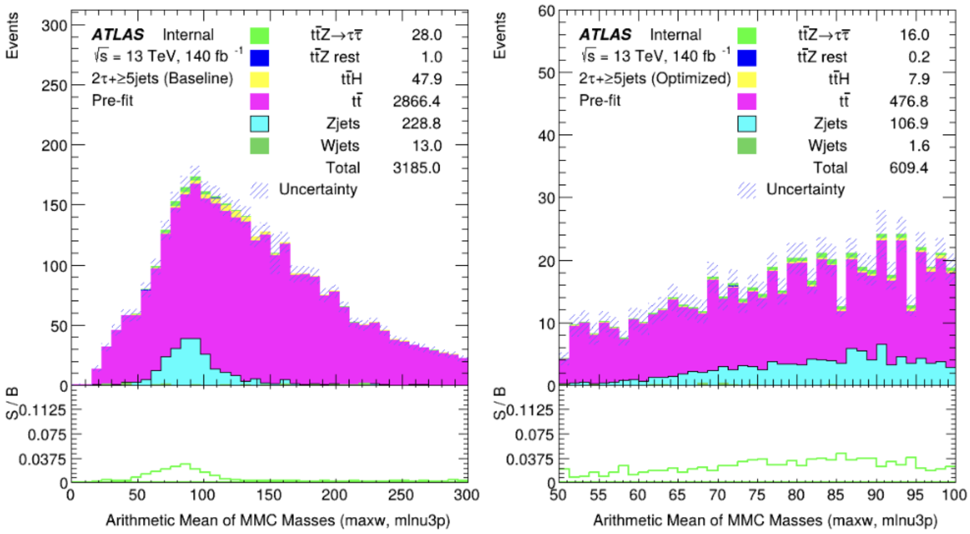

Figure 4: Impact of the MMC-based selection cut in the 0 lepton, 2 taus, and 5 jets region. The left plot shows the distribution of the arithmetic mean of the MMC mass estimates before applying the optimized selection, while the right plot shows the same distribution after applying the selection cut.

3 Results

3.1 Significance Results

We evaluated the signal significance under different scenarios, such as removing systematic uncertainties and increasing luminosity. The optimized region consistently outperformed the baseline across all tested configurations, which were selected based on the improved signal-to-background separation achieved in regions listed in Figure 3.

Despite optimization, systematic uncertainties remain a major limiting factor for discovery. For example, when scaling to 20× the current Run 2 luminosity (corresponding to approximately 3000 fb-1, as expected at the end of the High-Luminosity LHC era), the significance exceeds 5 standard deviations if systematics are fully removed. However, when reintroducing realistic 5% cross section uncertainty on the dominant $t\bar{t}$ background and 15% on the $t\bar{t}H$ contribution—the significance drops to approximately 2.2σ.

Regions Combined	Raw	Systematics Removed	Lumi ×20	Sys. Removed + Lumi ×20	Added Sys. + Lumi ×20
Baseline	1.13σ	1.23σ	2.71σ	5.52σ	2.04σ
Optimized	1.19σ	1.30σ	2.81σ	5.83σ	2.23σ

Table 2: Signal significance under different assumptions.

3.2 Luminosity Extrapolation

To estimate discovery potential at higher luminosities, we extrapolated the statistical significance using the standard relation (valid only when systematic uncertainties are negligible):

$$\text{Significance }(\sigma) \propto \sqrt{L}$$

where L is the integrated luminosity. Figure 5 shows the signal significance as a function of luminosity, assuming the same selection efficiencies and background/systematic uncertainties as in our current analysis. Our results suggest that the five σ discovery threshold cannot be reached with the current Run 2 dataset ($140\,\text{fb}^{-1}$). However, the High-Luminosity LHC (HL-LHC), with a projected $3000\,\text{fb}^{-1}$, offers potential for observing this rare decay channel.

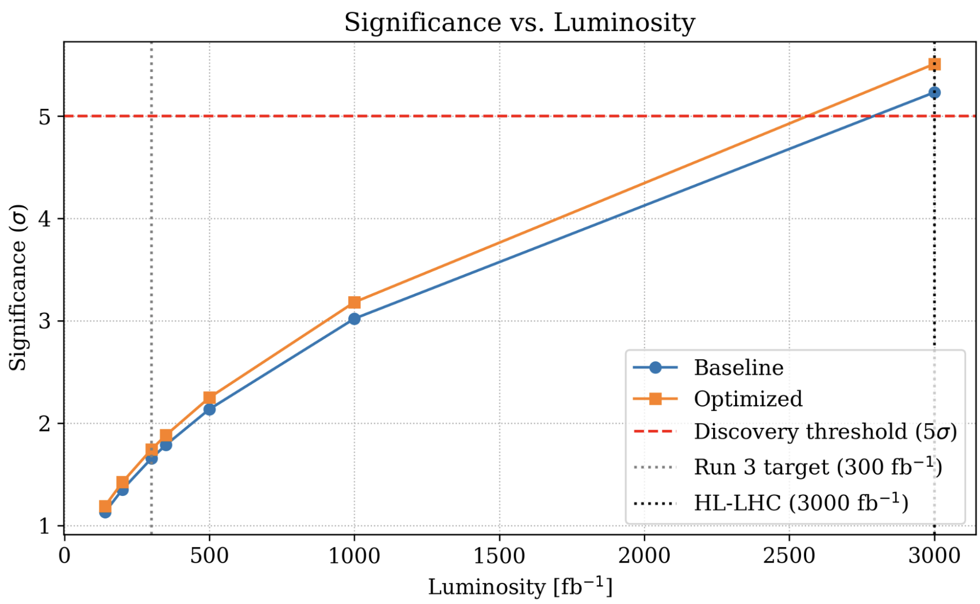

Figure 5: Projected signal significance under increasing integrated luminosity. The plot compares baseline and optimized selections. A five σ discovery threshold and target luminosities for Run 3 and HL-LHC are indicated.

3.3 Outlook: Machine Learning for Signal-Background Separation

After completing the cut-based analysis, we began exploring machine learning techniques, specifically deep neural networks (DNNs), to improve signal–background separation. The DNN approach allows for multivariable correlations that are difficult to capture with traditional cut-based methods.

We aim to revisit regions that showed limited separation in our manual analysis and evaluate whether a trained DNN can better distinguish signal from background events. This approach could be especially valuable in regions with subtle kinematic differences or overlapping features. The training uses a diverse set of kinematic inputs, including basic observables such as jet, tau, and lepton kinematics, as well as more complex variables like top quark mass and p_T (Top1, Top2), ΔR between reconstructed tops, and MMC-based outputs such as (`maxw`) and MET-related features.

Among all the input features, the reconstructed mass of the leading top quark emerged as the most influential in the classifier's performance, followed by the mass of the second top quark, the leading tau p_T, and the leading jet p_T. These insights suggest that variables encoding overall event topology and heavy-object reconstruction play a central role in signal discrimination.

After training a simple binary classifier DNN using these variables, we observed a clear improvement: the DNN increased the signal significance in this region to 0.58σ. This demonstrates the potential of even a minimal DNN model to recover sensitivity in challenging regions where manual cuts previously yielded no improvement (e.g., 0.31σ before ML).

The model was trained using $t\bar{t}Z(\tau\tau)$ events as the signal class and $t\bar{t}$ as background. The input features included transverse momenta, masses, and angular separations of final-state particles, with appropriate zero-padding for missing objects. The training and testing curves remained stable across folds, with AUC scores ranging between 0.75 and 0.77.

Figure 6 shows the confusion matrix between the true and predicted classes. The output distribution from the DNN, shown in Figure 8, confirms the separation between signal and background. The ROC curve in Figure 7 further quantifies the classifier's performance, and the final output-based separation achieved was 19.8%, providing additional justification for incorporating machine learning in future iterations of this analysis.

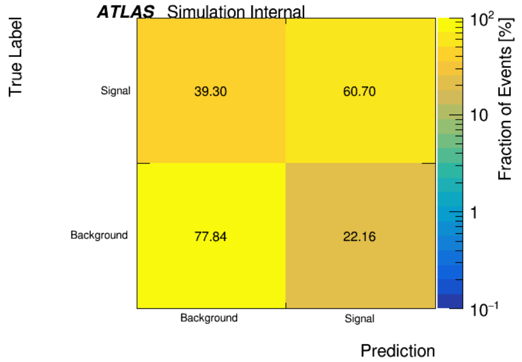

Figure 6: Confusion matrix for DNN classification of $t\bar{t}Z(\tau\tau)$ (signal) vs. $t\bar{t}$ (background).

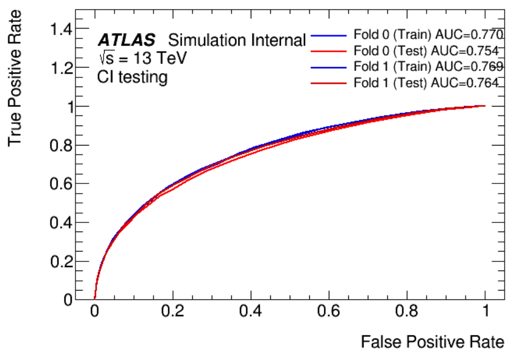

Figure 7: ROC curves for training and test folds showing AUC performance.

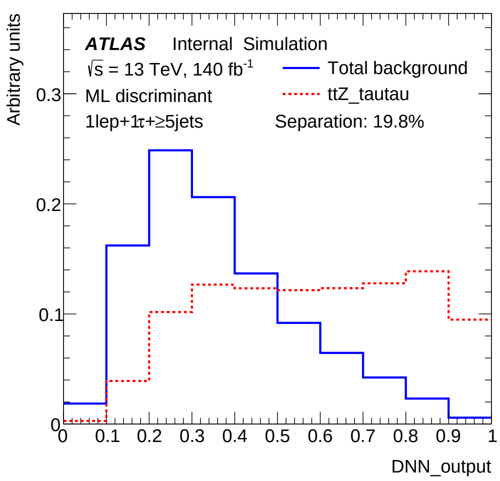

Figure 8: Output distribution of the DNN discriminant for signal and background in region `lep1_1tau_5jet`.

4 Conclusion

We investigated the feasibility of measuring the rare process $t\bar{t}Z(\tau\tau)$ using the full Run 2 ATLAS dataset. Through region design, kinematic variable exploration, and cut optimization, we improved signal-to-background separation and quantified the resulting significance. The optimized region selections consistently outperformed the baseline, though systematic uncertainties—especially from dominant backgrounds like $t\bar{t}$ and $t\bar{t}H$—remain a key limiting factor for discovery.

Significance extrapolation showed that the 5σ discovery threshold is unattainable at current integrated luminosity, but may be achievable at the HL-LHC (3000 fb^{-1}), provided systematic control is improved.

Finally, we initiated a machine learning–based approach using deep neural networks to enhance sensitivity in complex or challenging regions. Even a simple DNN model was able to recover signal significance in regions where manual methods failed, boosting the significance from 0.31 to 0.58. These results point toward the strong potential of ML techniques for future iterations of this analysis and other rare decay searches.

Together, these efforts lay the groundwork for the eventual measurement of $t\bar{t}Z(\tau\tau)$ and highlight the power of combining traditional and modern approaches in collider data analysis.

References

[1] ATLAS Collaboration. Inclusive and differential cross-section measurements of $t\bar{t}Z$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, including EFT and spin-correlation interpretations. *JHEP* **07** (2024) 163, https://arxiv.org/abs/2312.04450arXiv:2312.04450.

[2] ATLAS Collaboration. Performance of Missing Mass Calculator in the $H \to \tau\tau$ analysis. ATL-PHYS-PUB-2015-026. `https://cds.cern.ch/record/2041461`

[3] A. Elagin, P. Murat, A. Pranko, and A. Safonov. A new mass reconstruction technique for resonances decaying to $\tau\tau$. *Nucl. Instrum. Meth. A* **654** (2011) 481–489, https://arxiv.org/abs/1012.4686arXiv:1012.4686.