CERN Summer Student Programme 2024 Project Report: Analysis of $H \to ZZ$ and $H \to bb$ decay channels in $\sqrt{s} = 13$ TeV ATLAS Open Data for Education and Outreach

Chang Hexiang

1 Introduction

In 2024, ATLAS plans to have a new release for the ATLAS Open Data for Education and Research program [1]. This program aims to make LHC experimental resources publicly accessible, providing documentation and explanations to users outside the CERN collaboration. The new 13 TeV release includes the 2015 and 2016 ATLAS datasets, bringing the total integrated luminosity up to 36.6 fb^{-1}. In the Education and Outreach sector, users can make use of existing Python notebooks which demonstrate data analysis of the Open Data for various experiments, such as the Higgs detection through $H \to ZZ$ or $H \to \gamma\gamma$. This student project aims to improve upon existing documentation and programming tools by building on the pedagogical explanations, as well as creating new Python tutorials using the data from the new release.

The project consisted of two main tasks.

1. Improve the clarity of the existing $H \to ZZ$ notebook, adding explanations to the physics and the data analysis. This notebook will then be ported to the new 13 TeV data release.
2. Perform a new analysis on the 13 TeV data for $H \to b\bar{b}$ decay channel.

2 Significance and Pedagogical Methodology

As this project is aimed at improving and creating Python notebooks meant for outreach and education, it is important to highlight the methodology in creating these tools. One of the main goals of this project was to make these Open Data notebooks as accessible as possible to all users, especially those inexperienced in particle physics and data analysis. As such, a large emphasis was placed on writing clear and understandable code, all while including explanations on all relevant physics and programming concepts.

The first task of the project was to refine and improve upon an existing $H \to ZZ$ decay analysis Python notebook, available on the ATLAS Open Data server. In its original state, the notebook could fully run the analysis, successfully outputting the final plot of the invariant masses of the decay products. However, it was written without sufficient explanatory comments, and it lacked exposition on particle physics theory, which was the theoretical backbone of the analysis. Such explanations are of utmost importance in this project, as users may not be completely familiar with the field and its intricacies. This methodology is also carried to the final task of the project, in which a brand new analysis notebook was made from scratch, for the $H \to b\bar{b}$ channel. In this task, the analysis and the according explanations were completed in parallel. Again, emphasis was placed on readable code and clear physics explanations. With this methodology focused on pedagogy, one can be assured that users can benefit maximally from these notebooks.

3 Higgs particle production and decay

The Higgs boson has four main production modes. The Feynman diagrams for these production modes are shown in Figure 1.

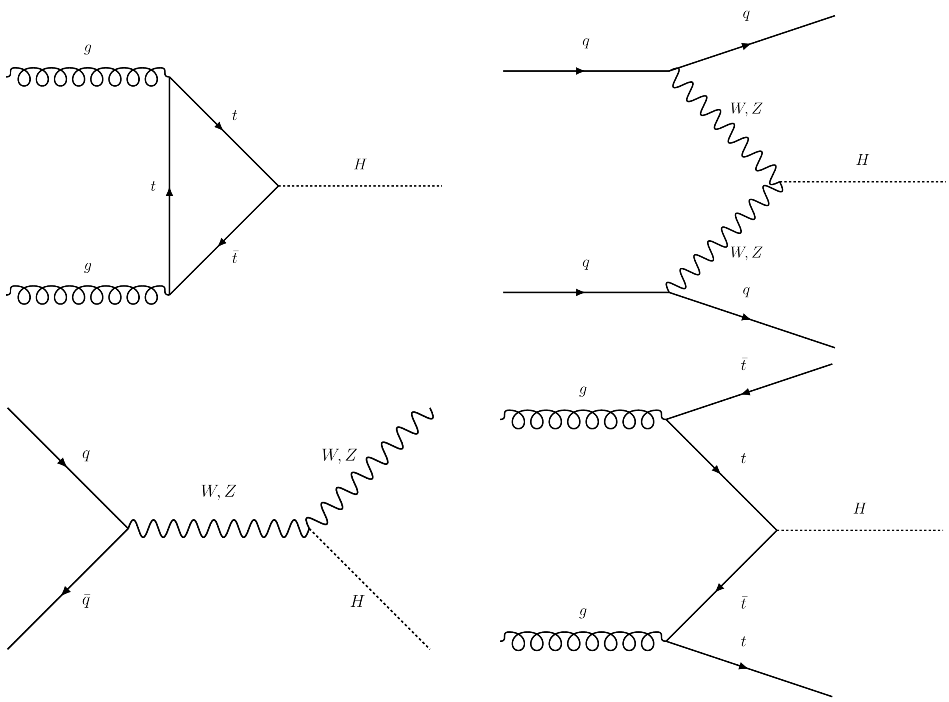

Figure 1: Higgs main production modes

In order of decreasing cross section, they are:

1. Gluon-gluon fusion (top left)
2. Vector boson fusion (top right)
3. Vector boson bremsstrahlung (bottom left)
4. Top-antitop fusion (bottom right)

The Higgs decays, on the other hand, take place in many channels. Each decay analysis will focus on one possible decay channel.

4 $H \to ZZ$ decay analysis

In the first task of this project, the $H \to ZZ$ decay is analysed. This decay was one of the first channels to be used to discover the Higgs boson in 2012 [5]. While the cross section of this decay is not the highest, the analysis is cleaner compared to other decay channels as the amount of background of this decay is relatively low, leading to a stronger signal significance.

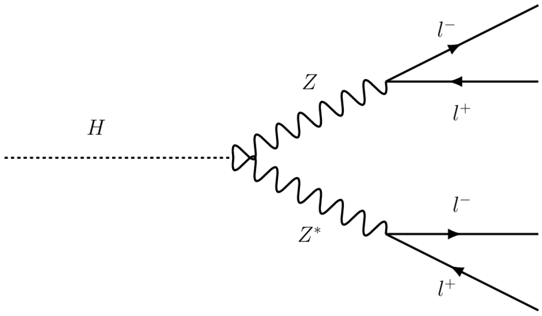

Figure 2: Higgs particle decay signal $H \to ZZ^* \to \ell\ell\ell\ell$

More specifically, we study the case of decay to four leptons, $H \to ZZ^* \to \ell\ell\ell\ell$, shown in Figure 2. We refer to this as our desired signal. Ideally, we would search for collisions which yield four leptons as products and this would tell us whether a Higgs boson is present. Unfortunately, in addition to our signal, there are many other background processes that lead to four reconstructed leptons in the final state. The main background is $ZZ^* \to \ell\ell\ell\ell$, where decay products have the same properties as those in the Higgs decay. This is known as an irreducible background. This process is shown in Figure 3.

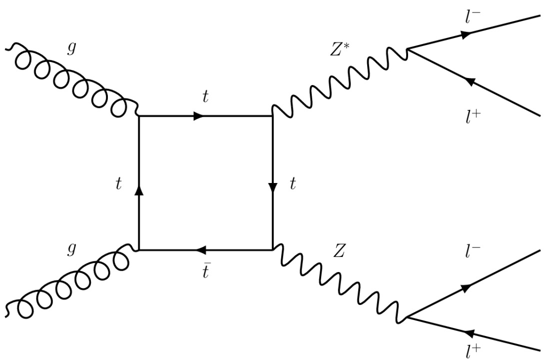

Figure 3: Irreducible $ZZ^* \to 4\ell$ background against the Higgs signal.

We can get around this by accounting for the total invariant mass of the lepton products. We know through conservation of energy and momentum that the invariant mass of the products must be equal to the Higgs mass, while other background processes will have different invariant masses. Our last step would be to plot the invariant mass of each event and spot the peak in mass around 125 GeV, which corresponds to the mass of the Higgs boson.

We also have background contributions from Z+jets and top-antitop processes, where additional charged leptons can arise either from semi-leptonic decays of heavy flavour or light flavour jets misidentified as leptons. These are shown in Figure 4. These backgrounds are difficult to remove completely. For such processes, we will use some selections to distinguish them from the Higgs decay.

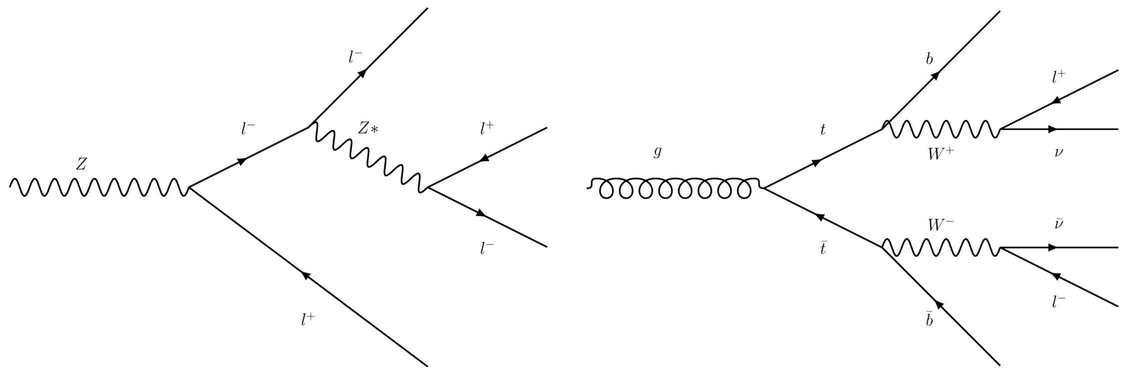

Figure 4: Reducible backgrounds $ZZ^* \to 4\ell$ (left) and top quark processes (right).

Selections

Using the properties of the leptons, we can generate some selection rules to cut away our background and purify the signal. Because the Higgs boson is a neutral particle with zero lepton number, the lepton products from its decay must sum to zero charge and zero lepton numbers. Thus, we can cut away all data with decay products that do not have these properties. Additionally, we introduce the following cuts to further improve the signal-to-background by cutting away regions that have large backgrounds but small signals.

1. leading lepton $p_t > 30\,\mathrm{GeV}$
2. subleading lepton $p_t > 20\,\mathrm{GeV}$
3. third-leading lepton $p_t > 10\,\mathrm{GeV}$

These cuts increase the ratio of our signal to the reducible background.

Result

We plot the total invariant masses of the lepton products in Figure 5. We identify the peak in the invariant mass spectrum at 125 GeV as a detection of the Higgs boson. The notebook is available on the ATLAS Open Data Github repository [2].

Porting the notebook

Once the analysis was completed and refined, the contents of the notebook were ported to include data from 2015 and 2016 of the new open data release. However, due to certain issues in the ongoing production of the datasets, the porting process was left incomplete. Nevertheless, the analysis and explanations remain identical to that of the old release.

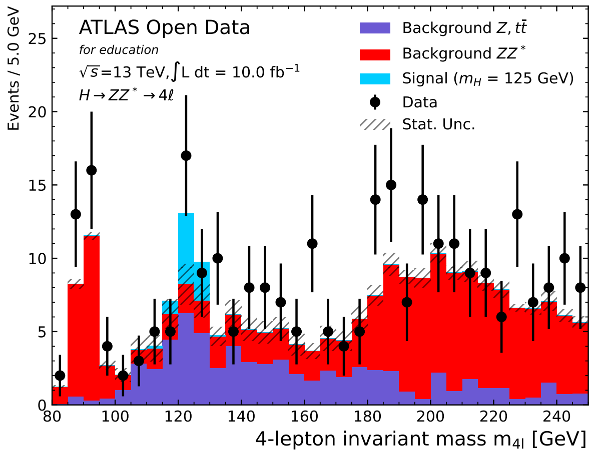

Figure 5: Histogram plot of the invariant masses of the four-lepton products

5 $H \to b\bar{b}$ decay analysis

In the final task, a new analysis was started on the decay channel $H \to b\bar{b}$. This decay is shown in Figure 6. This decay channel actually accounts for almost 60% of all Higgs decays.

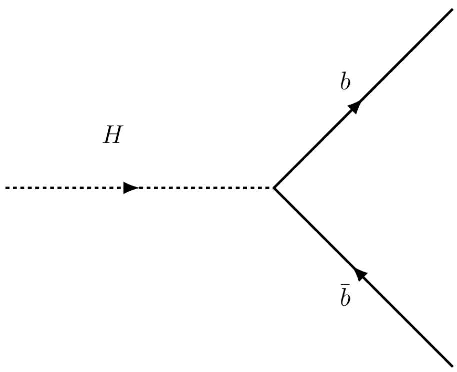

Figure 6: Higgs particle decay signal for $H \to b\bar{b}$

Ideally, we would search for collisions which yield two bottom quarks as products and this would tell us that a Higgs boson is present. Despite this, the Higgs was not originally discovered through this channel, and this decay was only observed more recently in 2018 [4]. In fact, this channel is one of the most difficult to detect at the LHC, precisely because proton-proton collisions tend to produce a huge number of b quarks independently of the Higgs. More precisely, in addition to our desired decay signal $H \to b\bar{b}$, there are many other background processes that lead to two reconstructed b quarks in the final state.

To make the analysis easier, we focus on the vector boson bremsstrahlung production channel, in which a W or Z bozon is also created. Because the leptonic decays of the W and Z have very clean signatures, these processes will help us identify the case where the Higgs is produced.

There are three main decay channels where $H \to b\bar{b}$.

1. 0-lepton products: $Z + H \to \nu\bar{\nu} + b\bar{b}$
2. 1-lepton products: $W + H \to l\nu + b\bar{b}$
3. 2-lepton products: $Z + H \to l\bar{l} + b\bar{b}$

In this analysis, we focus on the 0-lepton channel. This is referred to as our desired signal. In this channel, there are no charged leptons produced, only neutrinos and bottom quarks. Nevertheless, after narrowing our analysis, we still have to deal with a lot of background processes.

It must be noted that the ATLAS detector cannot detect neutrinos. Neutrino detection is presumably important, as we expect two neutrino products in the 0-lepton channel. While direct detection is not possible with ATLAS, conservation of momentum allows us to indirectly detect the presence of neutrinos. Since the proton beams in the LHC are travelling axially, the transverse momentum must sum to zero before and after the collision event. When a neutrino is created and not directly detected, there is some missing transverse energy E_T^{miss} that can be calculated from the momenta of the other decay products. This missing transverse energy can then be used in calculations to verify the Higgs decay.

One of the main sources of background is that of V+jets processes, where leptons produced can either be lost or misidentified as jets. One possible Feynman diagram is shown in Figure 7.

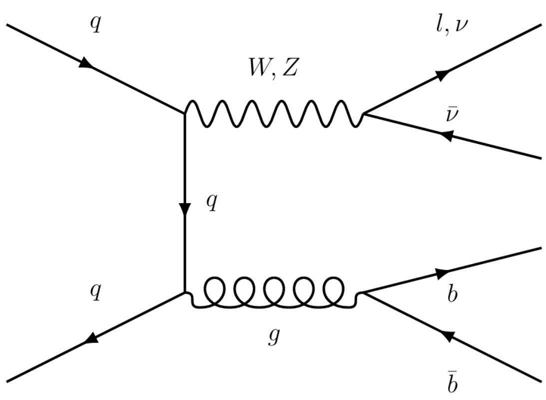

Figure 7: Main background V+jets process

We also have background contributions from the diboson processs $ZZ \to \nu\bar{\nu} + b\bar{b}$ and top-antitop processes, shown in Figure 8.

For these backgrounds, we will again distinguish them from the Higgs decay using various selection rules. Due to the incredibly small signal-to-background ratio of this decay channel, we require a large number of selections.

Selections

Our goal is to isolate the signal from our background. These selection rules place restrictions on the physical variables in the data, such as the relationships between the jet and

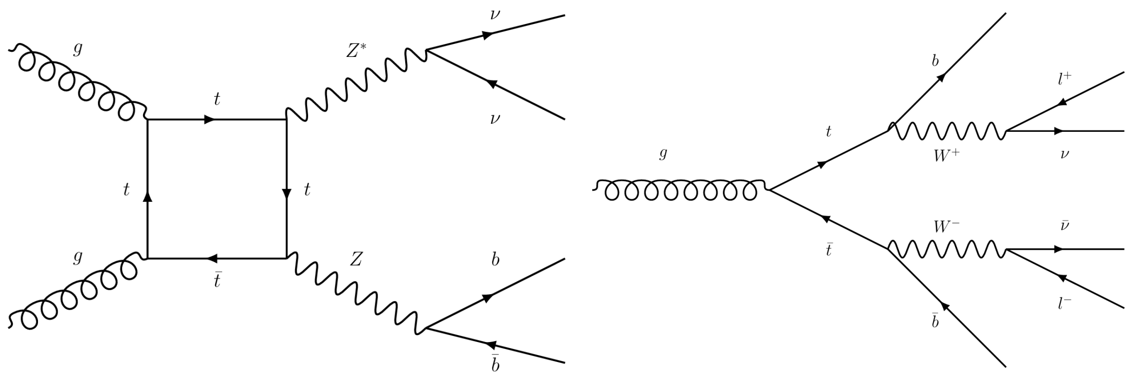

Figure 8: Backgrounds $ZZ^* \to \nu\bar{\nu} + b\bar{b}$ (left) and top quark processes (right).

neutrino angles, or the minimum E_T^{miss}. In its current state, the data contains many unwanted detections and false positives. There are many reasons for this. For one, it is possible to misidentify leptons as jets. Some of our selections will aim to cut away such misidentifications based on the properties of the jets. We make the following selections on the data.

Missing Transverse Energy (MET): Passes the online MET trigger. Also, offline we require $E_T^{\text{miss}} > 150\,\text{GeV}$.

- Trigger selection cuts away invalid detections.

Lepton cuts: Events must not contain any loose leptons.

- A set of cuts that determine the level of efficiency of a lepton identification algorithm.
- Looseness and tightness of lepton detections define a strictness criteria for whether a lepton was truly detected.
- In the selection, loose leptons are defined via the inverse of the tightness ID, and also have $E_T > 25\,\text{GeV}$.

Jet cuts: Exactly two or three jets. All jets must have $p_T > 20\,\text{GeV}$ and $|\eta| < 2.5$.

- Since the Higgs decays to two b quarks, we expect at least two jets in the desired signal. We shall restrict our selection up to three jets, rejecting events with four or more jets. This will reduce the large background arising from $t\bar{t}$ production.
- We restrict the value of pseudorapidity η because this is the accessible range of the inner detector which we use for b-tagging.

b-tagging: Exactly two b-tagged jets. For the b-tagged jet with highest p_T, we require $p_T > 45\,\text{GeV}$.

- A b-tagged jet is simply a jet that has been identified to be originating from a b quark. Because there are two b quarks in the Higgs decay, we demand only events with two of such jets. In practice, a jet will be b-tagged if its b-tag quantile is equal to 2.
- We also place a lower bound cut on the p_T of the leading b quark, as this will significantly reduce the contribution of background processes, increasing the significance of the Higgs signal.

Sum of jet p_T, $H_T > 120\,\text{GeV}$

- This selection is used to remove a marginal region of phase space in which the trigger efficiency exhibits a small dependence on the jet multiplicity.

Angular selections: Require $\Delta\phi$ between two b-tagged jets $< 140°$, $\Delta\phi$ between E_T^{miss} and sum of b-tagged jets $> 120°$, and minimum $\Delta\phi$ between E_T^{miss} and various jets must be $> 20°$ for two-jet events and $> 30°$ for three-jet events.

- These angular selection criteria are applied in order to suppress multi-jet background. These backgrounds mostly consist of jets mismeasured in the calorimeters.

Result

We plot the total invariant masses of the two b quark products in Figure 9. The notebook is available on the ATLAS Open Data Github repository [3].

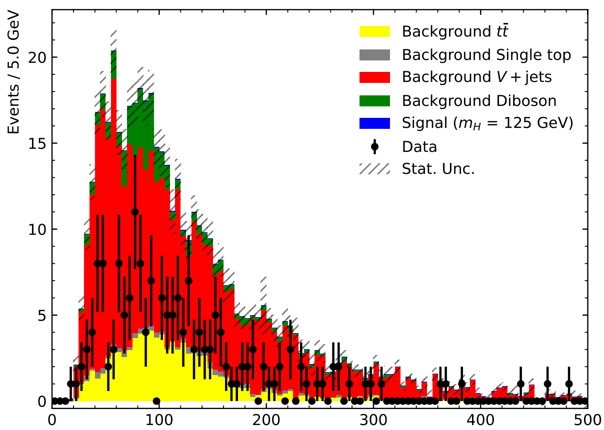

Figure 9: Histogram plot of the invariant masses of the two b quark products

6 Conclusion

In summary, this project successfully refined the existing $H \to ZZ$ analysis notebook and partially ported it to the new release. This serves as a useful learning tool for any users with interest in learning about particle physics and data analysis. Additionally, the $H \to b\bar{b}$ analysis has mostly been completed with full documentation. However, it remains a work in progress as the analysis needs to be verified and the underlying datasets need to be refined further. Contributors to ATLAS Open Data are encouraged to bring this analysis to full completion.

References

[1] ATLAS Open Data: High Energy Physics data for everyone. `https://opendata.atlas.cern/`, 2024.

[2] The ATLAS collaboration. How to rediscover the Higgs boson yourself! https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/uproot_python/HZZAnalysis.ipynb, 2024. Accessed on 5 September 2024.

[3] The ATLAS collaboration. Looking for the Beauty Higgs decay! https://github.com/atlas-outreach-data-tools/notebooks-collection-opendata/blob/Htobb/13-TeV-examples/uproot_python/HbbAnalysis.ipynb, 2024. Accessed on 5 September 2024.

[4] The ATLAS collaboration., Aaboud, M., Aad, G. et al. Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector. *J. High Energ. Phys.*, 24(2017), 2017.

[5] The ATLAS collaboration., Aaboud, M., Aad, G. et al. Measurement of the Higgs boson coupling properties in the $H \to ZZ \to 4\ell$ decay channel at TeV with the ATLAS detector. *J. High Energ. Phys.*, 95(2018), 2018.