Measuring the Mass of Z boson and Determining the Existence of the Higgs Boson using ATLAS Open Data

Zayd Pandit (PNDZAY001), PHY3004W
Due: 24 August 2021

***Abstract*—Using data from the ATLAS experiment at CERN provided through ATLAS Open Data [cite], we will measure the mass of the Z boson by observing a proton collision and analysing the $pp \to Z \to \mu\mu$ decay. We also investigate a potential $H \to \gamma\gamma$ by analysing its products, and estimate the statistical significance of the existence of the Higgs Boson in this decay.**

I. Introduction and Theory

THE ATLAS detector is the largest general-purpose detector at CERN. The ATLAS Open Data project provides data from the ATLAS detector to the public for educational and research purposes. We will be using the 13TeV proton-proton collision dataset, with an integrated luminosity of $10\, fb^{-1}$.

During a proton proton collision, a quark and an antiquark can interact via the weak force, producing the Z-boson. The Z-boson has a large mass of approximately $92000 MeV$, so it decays very quickly into various particles in a variety of decay chains. The easiest decay to reliably detect is the decay into muon pairs ($Z \to \mu^+\mu^-$). From the 4-vectors of the detected muons, we can use conservation of 4-momentum, to reconstruct the $\mu\mu$ system and thus find the 4-vector and mass of the Z-boson.

Since Z has a large mass the muons it decays to should have a large p_T and thus will be detected in the low η area of the detector [3]. We know that the muons from the Z decay will be more isolated in their cones compared to the muon products of common heavy hadron decay [3]. We can thus use this to eliminate detections which contains a large total sum of energy in the surrounding area of detection.

The Brout-Englert-Higgs boson (Higgs Boson, H) is a massive chargeless, spinless boson theorized by Brout, Englert, Higgs, Guralnik, Hagen and Kibble in 1964. This existence of this Boson has been experimentally verified by the ATLAS experiment at CERN in 2012 . The Higgs Boson decays into two photons ($H \to \gamma\gamma$), with $m_{\gamma\gamma} \approx 125$ MeV [1]. In order to eliminate events that mimic photons, the data used in this investigation has been filtered such that the events are isolated from other particles and have a fairly high certainty. In order to determine if a signal from the decay of the Higgs Boson is present in the data, we plot the number of events minus the background as a function of their respective energies. If a signal is present that can be attributed to the Higgs Boson, we can quantify the likelihood that the data fluctuation is in fact a particle, rather than background events by defining the measure of statistical significance:

$$S = \frac{N_S}{\sigma(N_B)} \tag{1}$$

where N_S is the difference between the number of counts and the background ($N_S = N - N_B$) and $\sigma(N_B)$ is the total uncertainty of the number of background counts.

II. Results

A. Part 1: Z Boson Mass

The Jupyter Notebook provided (see Appendix 1) performs a basic analysis of the data. A few basic data cuts are applied - restrict to large transverse momentum, restrict to small pseudorapidity. The events were histogrammed into 250 bins, and fitted to the Breit-Wigner distribution by the method of Chi-square minimization. The method of chi-square minimization used here just varied the mass parameter and did not optimize the characteristic width or the scaling factor of the Breit-Wigner distribution. The algorithm employed also only allowed for integer values for the mass - another major weakness. This initial analysis gave a mass for the Z-boson of:

$$m_Z = 90764.00000 \pm 7.35222 \text{ MeV} \tag{2}$$

The first change to the initial analysis was to optimize the cuts. This was achieved through creating arrays of possible cut-off values. The entire analysis was then iterated through, while varying the value of a single cut parameter. The value which results in the minimum uncertainty is then selected as the optimal parameter value - if a clear minimum was present. The first cut was to restrict muon pair detections to only those which have opposite charge. The second cut was to restrict the pseudorapidity of both muons to be less than 2.42. The third cut restricted (etcone/pt) to be less than 0.1 for each muon, where etcone is the scalar sum of track E_T in a cone of R=0.2 around the muon detection, and pt is the muon's transverse momentum.

There are various statistical distributions which can be used to fit the data. The events should theoretically be distributed according to Breit-Wigner (or Cauchy) distribution [2]. However, this distribution is known to not be able to account for deviation due to detector resolution or relativistic phenomena. However, the Relativistic Breit-Wigner can better account for the relativistic effects during the high speed collisions. A double Gaussian, with their mean values constrained to be equal, can also be used to fit the data well. This is due to the large number of events and the central limit theorem meaning that radiative effects, background events and other largely random variables will be accounted for better than in the Breit-Wigner type distributions.

A key refinement to the analysis was to only use the data around the peak signal. Much of the other data will be background and due to other processes, so by isolating the peak, we effectively cut the data to signals which are more likely from the $Z \rightarrow \mu^- \mu^+$ decay. We can notice the importance of this cut by noting the lack of agreement between the fitted distributions at the tails, and by the improvement of the reduced chi-square statistic (χ^2_ν). The χ^2_ν value decreased from 130.45 to 6.25 when using the Breit-Wigner fit with the basic χ^2_ν minimization technique.

The binning of the data can affect the quality of the fit to the model distribution. Increasing the number of bins tended to improve the χ^2_ν value, up to a certain point. Having too many bins results in there being too little events in many bins for their Poisson distributed uncertainties to be accurate (an arbitrary bin with no counts should still have a reasonable uncertainty). We used 1000 bins in our final analyses.

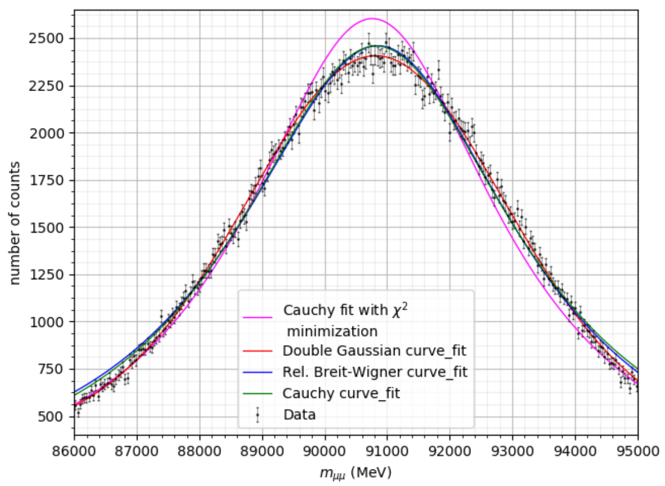

Fig. 1. The optimized fits of the Cauchy, Double Gaussian, Relativistic Breit-Wigner distributions, where the Cauchy distribution was fitted with χ^2 minimization and curve_fit, and the rest by curve_fit

TABLE I
MEASURED MASS OF THE Z-BOSON ALONG WITH ITS ASSOCIATED UNCERTAINTIES AND THE REDUCED CHI-SQUARE OF THE FIT UTILISED

Fitting routine	m_Z (MeV)	χ^2_ν
Breit-Wigner by χ^2 minimization	90788.00000 ± 7.95222	10.4899
Breit-Wigner by curve_fit	90825.71671 ± 8.61556	2.48220
Double Gaussian by curve_fit	90818.52969 ± 6.09141	1.08146
Relativistic Breit-Wigner by curve_fit	90855.05428 ± 8.68776	2.57457
Particle Data Group [cite]	91187.6 ± 2.1	-

B. Part 2: Detecting the Higgs Boson

A Jupyter Notebook was provided (see Appendix 1) including a basic analysis of the data for the possible $H \rightarrow \gamma\gamma$ decay. This included a cut to restrict the transverse momenta of the photon to above a threshold - since the predicted Higgs mass is really high, the photons it decays to should have large transverse momenta and low pseudorapidity. Using the same cut optimization process used in Part 1, the optimal cut was found to restrict photons to a transverse momentum over 59000. The photons pseudorapidity was restricted to be lower than a threshold - this was found to be optimal at pseudorapidity less than 1.37. The other cuts were kept as they were in the intial analysis. These being: Restrict pt/diphoton mass of one photon to be greater than 0.35, and the other to be greater than 0.25 (where pt is the transverse momentum of the photon). Restrict the diphoton mass to be between 105000 MeV and 160000 MeV

The bounds on which the distributions were fitted were restricted significantly to be between 120000 MeV and 131000 MeV. This change was of crucial importance and significantly improved the fit, and the stability of the fit upon change of variables such as number of bins.

Like in Part 1, more data bins results in a lower uncertainty of fit, but only up to a certain point. Unfortunately, my algorithm for optimization of bin number failed as there was no clear minimum of uncertainty present when varying the number of bins. The optimal value was approximated to be 10000 bins.

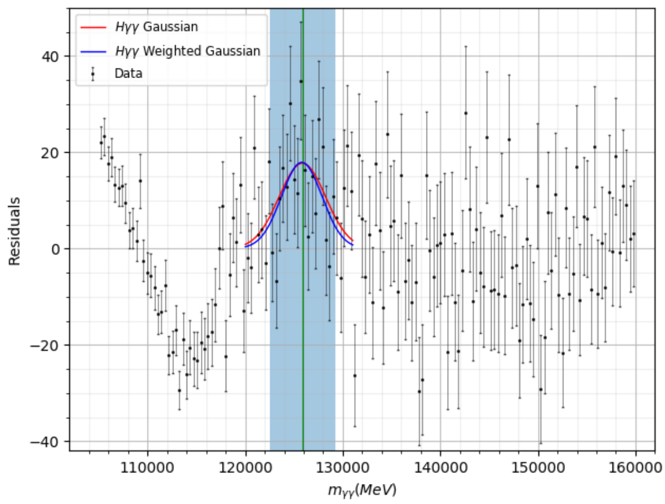

Fig. 2. A plot of the residuals of the number of events at each mass. The fitted Gaussian and weighted Gaussian, modelling the possible Higgs signal is overlayed with the data. The green vertical line shows the line at the m_H provided by the unweighted Gaussian fit, with an interval $\pm\sigma(m_H)$ from the mean shaded in blue.

TABLE II
MEASURED MASS OF THE Z-BOSON ALONG WITH ITS ASSOCIATED UNCERTAINTIES AND THE REDUCED CHI-SQUARE OF THE FIT UTILISED

Fitting routine	m_H (MeV)	$\sigma(m_H)$	S
Gaussian by curve_fit	125863.33055	3367.21985	3.08624
Weighted Gaussian by curve_fit	125805.48536	2936.86982	2.75231
Particle Data Group [1]	125100	140	≥ 5

III. Discussion

A. Part 1: Z Boson Mass

It is clear from Figure 1 that the double Gaussian fit our data the best - and the reduced chi square statistic (χ_ν^2) and the uncertainty given in Table 1 agree with this. The χ_ν^2 of the Double Gaussian is very close to 1 and shows a very accurate fit of the data. The other fits have values of χ_ν^2 which are all more than a factor of 2 bigger than the χ_ν^2 of the Double Gaussian fit.

In Table 1 we can see that our measured values disagree with each other within their respective uncertainties. This could indicate that our uncertainties were underestimated. A possible reason for this would be the optimizing of cut parameters to create lower uncertainties - this approach to reduce uncertainty could have resulted in an increase in precision, without necessarily improving the accuracy of the result.

It was noted during the investigation that the `scipy.optimize.curve_fit` consistently resulted in a similar uncertainty and a lower χ_ν^2 than the simple chi-square minimization procedure. This is exemplified through the comparison between the two Breit-Wigner fits done - The χ_ν^2 of the curve_fit was less than half that of the chi-square minimization, and there uncertainties are within 10% error of each other. The chi-square minimization procedure employed here, was a very rough, inaccurate method. This was due to the fact that it did not utilize the optimization of the scaling factor and characteristic width. This procedure also did not take into account the Poisson uncertainty on each bin of data. It is possible, perhaps in a follow up investigation, to enhance the chi-square minimization by weighting counts and minimizing the chi-square for all 3 parameters (mean, width, scaling factor) in multivariate version of this optimization scheme.

Due to the incredibly good fit of the double Gaussian, we can say with confidence that our measurement of the mass of the Z-boson using the 13TeV data set from ATLAS Open Data is $m_Z = 90818.52969 \pm 8.61556$ MeV. This does not agree with the incredibly well verified value of 91187.6 ± 2.1 MeV provided by the Particle Data Group (PDG) [4]. The difference cannot be accounted for by the uncertainties, due to their tiny magnitude.

The incongruency between our result and that published by the PDG can be due to a variety of factors. These include but are not limited to: The PDG used a different dataset (even though they observed the same collision in the same detector). Various experimental Type B uncertainties which were unaccounted for in the various data categories. We did not use a proper model of the background - there are known methods of modelling the background events using functions such as a decaying exponentials.

B. Part 2: Detecting the Higgs Boson

When fitting our residuals with an unweighted Gaussian, we find that the peak signal has a statistical significance of 3.08624. This sufficient to say that the signal we detected is the Higgs Boson and not random background fluctuations with a confidence of 99.73%. Our measurement thus fairly strongly confirms the existence of the Higgs Boson. The 5σ criteria used by the PDG has a corresponding p-value of 0.9999994. Thus the PDG measurement has a far better reliability than our measurement.

Our analysis is shown to be inaccurate to to the discrepancy in the Higgs mass measured by us and measured by the PDG. This difference in mass shows that our fit of the data could be inaccurate. Thus our statistical significance we measured could be misrepresenting the data. Further analysis of our data is definitely needed in order to reconcile the difference in measured Higgs mass. Further cuts that isolate events

IV. Conclusion

Our best measurement for the mass of the Z-boson is $m_Z = 90818.52969 \pm 8.61556$ MeV - this disagrees with the value of 91187.6 ± 2.1 MeV measured by the PDG [cite]. An improved analysis, along with additional datasets will help to reconcile this incongruency.

Our measurement of the a signal in the potential $H \to \gamma\gamma$ decay suggest fairly strongly that the signal is the Higgs Boson, confirming its existence. Our measurement of the Higgs signal produced a theoretical mass disagreeing with the value measured by the PDG, suggesting that here further analysis is needed too

V. Acknowledgements

I'd like to acknowledge the contributions of Faaris Alam and Victor Bantchovski, who assisted me in designing the python code for iteration through various cut parameters and evaluation of the resulting fit.

References

[1] P.A. Zyla; et al. *Higgs Boson. Progress of Theoretical and Experimental Physics (2020)*. URL: https://pdg.lbl.gov/2020/listings/rpp2020-list-higgs-boson.pdf.

[2] David Griffiths. *Introduction to Elementary Particles*. WILEY-VCH, 2008. ISBN: 9780198520115.

[3] James Keaveney. *Rediscover the Higgs Boson 2021*. University of Cape Town, 2021.

[4] M. Tanabashi. "Particle Data Group: Review of Particle Physics". In: *Physical Review* 98.3 (2018). DOI: http://pdglive.lbl.gov/Particle.action?node=S044.

VI. Appendix

1. The Jupyter Notebooks for the Z-Boson measurement and the Higgs Boson analysis can be found in the following repository: https://github.com/keaveney/UCT3rdYearLabATLASOpenData