A Higgs-like Particle of 125 GeV from $H \to ZZ \to 4l$ Search

Yongao Hu*
MIT Department of Physics
(Dated: April 8, 2025)

This report presents an excess in signal of a Higgs-like particle around 125 GeV in the $H \to ZZ \to 4l$ search. The data is collected from the 2011 and 2012 LHC run at $\sqrt{s} = 7$ TeV and 8 TeV respectively. The results are consistent with the Standard Model prediction of a Higgs-like particle with a mass of 125 GeV, with a significance of 2.33 σ.

I. INTRODUCTION

Higgs boson and the corresponding Higgs mechanism are the key components of the Standard Model of particle physics (SM). The Higgs boson is responsible for giving masses to SM particles. The discovery of the Higgs boson at the Large Hadron Collider (LHC) in 2012 is one of the most important achievements in particle physics in the past decade [8]. This paper aims to reproduce the H→ZZ search in [8] and analyze the data collected by CMS experiments from the 2011 and 2012 LHC run at $\sqrt{s} = 7$ TeV and 8 TeV respectively to search for a Higgs-like particle with a mass of 125 GeV.

II. THEORETICAL BACKGROUND

II.1. Higgs Bosons

Higgs boson, a spin-0 massive particle, is a key component of SM. The corresponding Higgs field is a scalar field that permeates all of spacetime, and its non-zero vacuum expectation value (vev) leads to electroweak symmetry breaking, which gives masses to the W and Z bosons [19, 22]. Chiral fermions such as the matter content in SM also acquire mass through their interaction with the Higgs field [19, 22]. Up until 2010s, Higgs boson has not been observed.
The Higgs boson is produced in high-energy collisions, such as those at the LHC [8]. The decay channels of the Higgs boson include H→ZZ, H→WW, H→bb, and H→$\tau\tau$ [12, 17].

II.2. H→ZZ Decay

The decay of the Higgs boson into two Z bosons (H→ZZ) is a rare process, with a branching ratio of about 2.5% for a Higgs boson mass of 125 GeV [17]. The decay products of the Z bosons can be reconstructed from their decay into four leptons (pairs of e^+e^- and/or $\mu^+\mu^-$). The final state consists of four charged leptons, which can be used to reconstruct the mass of the Higgs boson. The decay of the Z bosons into two charged leptons ($Z \to ll$) provides a clean signature for the Higgs boson search. This is because the products are non-hadronic, avoiding jet and hadronization modeling, and there are no neutrino produced, i.e. no missing signals. The mass of the Higgs boson can be reconstructed from the invariant mass of the four-lepton system.

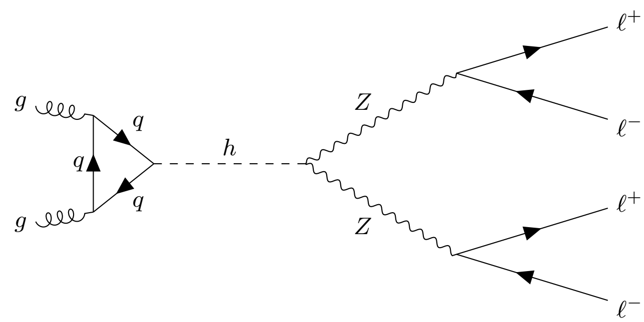


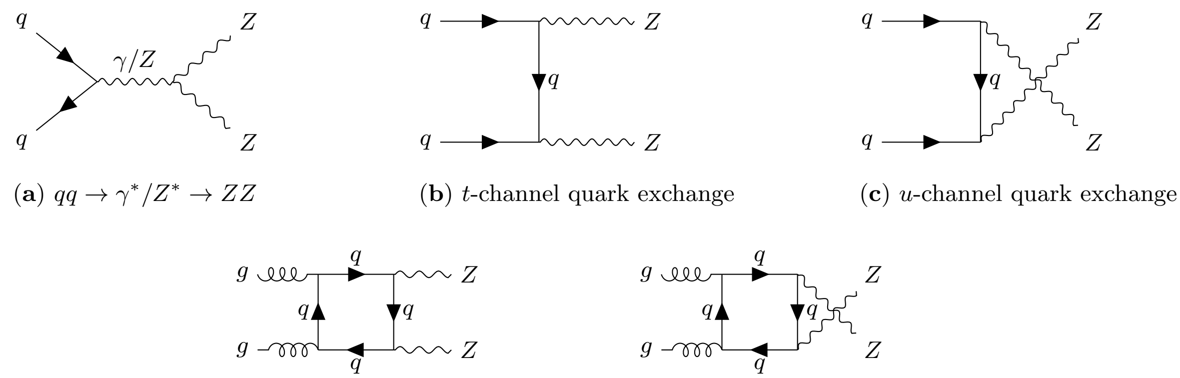
FIG. 1. Feynman diagram for the Higgs boson decay into two Z bosons and 4 leptons. The gluon fusion process is shown, where the Higgs boson is produced through a top or bottom quark loop. The gluons are produced when two protons collide with each other.

III. BACKGROUND EVENTS

Nevertheless, there are several background processes that are irreducible, meaning that they mimic the signal and cannot be completely removed by selection criteria. The largest background is the continuum ZZ^* production. Some of the production processes of ZZ^* are shown in Figure 2. The continuum ZZ^* production is a QCD process that produces a pair of Z bosons from the annihilation of quarks and antiquarks or gluon-gluon fusion.

Besides, there are additional backgrounds from the production from Drell-Yan process, a QCD process that produces a pair of leptons from the annihilation of quarks and antiquarks, and $t\bar{t}$ production, which is the pair production of top and anti-top quarks that decays into leptons. These processes can also produce events that are irreducible and resemble signals, but they are less significant than the continuum ZZ* production.

* yongao@mit.edu



(a) $qq \to \gamma^*/Z^* \to ZZ$ (b) t-channel quark exchange (c) u-channel quark exchange

(d) $gg \to ZZ$ box diagram (e) $gg \to ZZ$ box diagram (crossed)

FIG. 2. Representative Feynman diagrams for continuum ZZ^* production: (a)–(c) from quark–antiquark initial state, (d)–(e) from gluon–gluon initial state.

IV. EXPERIMENTAL BACKGROUND

LHC is a proton-proton collider located at CERN, Geneva, Switzerland [8]. The LHC is the largest and most powerful particle accelerator in the world, with a design energy of 14 TeV. The CMS experiment is a general-purpose detector designed to study a wide range of physics processes, including Higgs boson searches. Here, we are using the Run 1 data with center-of-mass energy $\sqrt{s} = 7$ TeV (in 2011) and 8 TeV (in 2012) [8, 9, 13, 16].

IV.1. CMS Detector

CMS experiment includes a superconducting solenoid that provides a magnetic field of 3.8 T to bend the trajectories of charged particles, a silicon tracker to measure the momenta of charged particles, a lead tungstate crystal electromagnetic calorimeter (ECAL) to measure the energy of photons and electrons, a brass/scintillator hadron calorimeter (HCAL) to measure the energy of hadrons, and gas-ionization muon detectors [8].
Pseudorapidity η is used to describe the angular distribution of particles in the CMS detector. Pseudorapidity is defined as $\eta = -\ln(\tan(\frac{\theta}{2}))$, where θ is the polar angle from the positive z-axis. The silicon tracker tracks charged particles within the pseudorapidity range of $|\eta| < 2.5$. The ECAL and HCAL measures particles within the pseudorapidity range of $|\eta| < 3.0$. The muon detectors detect muons within the pseudorapidity range of $|\eta| < 2.4$ [8].

IV.2. Monte Carlo Simulation

The signal and background processes are generated using monte carlo simulation. First, the underlying hard-scattering processes are generated with POWHEG [4] or MadGraph4 [5], followed by parton showering and hadronization using PYTHIA6.4 [21]. The resulting events are then passed through a GEANT4-based [2] simulation of the detector. The parton distribution functions (PDFs) are chosen according to PDF4LHC working group recommendations [3, 6, 7, 15]. The simulation data utilized in this study is obtained from CMS Open Data [9, 13, 16].

V. SELECTION CRITERIA

V.1. Preliminary Selection

The data from [13, 16] has been pre-processed to remove the events that do not pass the preliminary selection criteria. The events are selected based on the following criteria [13]:

- The events contains only four leptons, which are either electrons or muons as the final state particles.
- Transverse impact parameter with respect to the primary vertex $|d_{xy}| < 0.5$ cm.
- Longitudinal impact parameter with respect to primary vertex $|d_z| < 1$ cm.
- 3D impact parameter significance: $|\text{SIP}| < 4$, where $\text{SIP} = \frac{I}{\sigma_I}$, I is the 3D lepton impact parameter, and σ_I is its uncertainty. This ensures that the lepton pairs from Z boson decays originate from the same primary vertex.
- Muon and electron selection criteria: relative isolation of the lepton (reIso), the scalar sum of the transverse momenta of particles reconstructed within a distance ΔR of the object, normalized to the p_T of the object, is smaller than 0.4 within a cone $\Delta R = 0.4$, where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. $\Delta\eta$ is the difference in pseudorapidity, and $\Delta\phi$ is the difference in azimuthal angle.

We further apply the following selection criteria to select the events that pass the preliminary selection [8]:

- Pseudorapidity of the lepton: $|\eta_e| < 2.5$ for electrons, $|\eta_\mu| < 2.4$ for muons.
- Transverse momentum of the lepton: $p_T > 20$ GeV for electrons, $p_T > 10$ GeV for muons.
- The four-lepton system includes two pairs of same-flavor opposite-charged leptons are required to form the four-lepton system.
- The pair of leptons with invariant mass closer to mass of Z boson have mass between 40 GeV and 120 GeV, and the other pair have mass between 12 GeV and 120 GeV.

V.2. Neural Network

The neural network (NN) is used to classify the events into signal and background, trained using the pre-processed data. The NN utilizes feed-forward multi-layer perceptron (MLP) [20] with PyTorch backend [18]. The input features of the NN are the x and y components of momenta, and particle ID of the four-lepton system of the Monte Carlo simulation with invariant mass between 100 GeV and 160 GeV. We did not use the z component of momenta or the mass of the particles because otherwise the NN can reconstruct the invariant mass and use that as the primary feature.

We use 72% of the data for training, 18% of the data for validation, and 10% of the data for testing. Each lepton is processed individually with three layers of size 32 each. Once the 4 leptons are combined, we have more fully-connected layers ($256 \to 64 \to 16 \to 4$) before the final output. The output is yes/no for signal/background.

The NN is trained using the Adam optimizer [14] with a learning rate of 3×10^{-4}. The loss function is binary cross-entropy [11]. The NN is trained for 100 epochs. The NN is trained to minimize the loss function, and the accuracy is calculated as the number of correct predictions divided by the total number of predictions. The training and validation loss and validation accuracy are shown in Figure 3.

V.3. Result

After applying the NN and the selection criteria to the monte-carlo simulation and data, we obtain the final dataset, shown in Figure 4. The dataset is used to calculate the significance of the excess in signal as a function of mass.

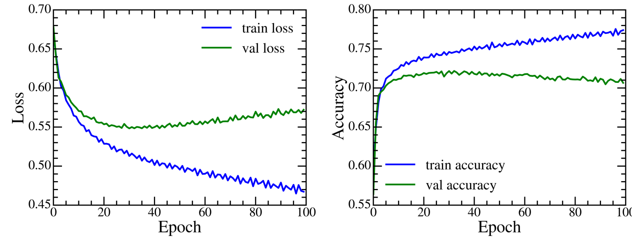


FIG. 3. Training and validation loss and accuracy of the neural network. The loss is calculated using binary cross-entropy. The accuracy is calculated as the number of correct predictions divided by the total number of predictions.

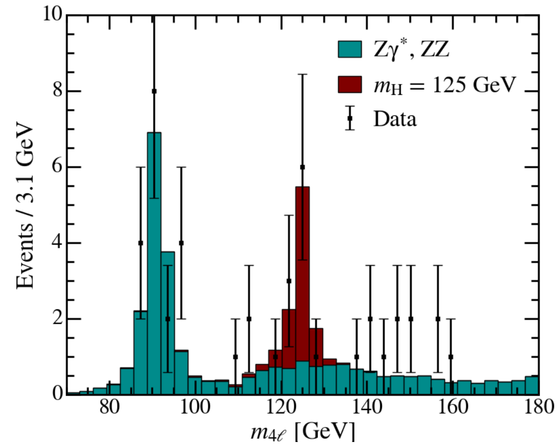


FIG. 4. Final dataset after applying the selection criteria and the neural network. The error bar comes from the standard statistical uncertainty of the data.

VI. STATISTICAL ANALYSIS

The statistical analysis is performed using the profile likelihood ratio test [10]. We have two hypotheses: the null hypothesis (H0) and the alternative hypothesis (H1). The null hypothesis is that there is no signal (background only), and the alternative hypothesis is that there is a signal (signal and background). We approximate the signal as a Gaussian distribution on top of the background.

We can calculate the p-value of the data using the likelihood ratio test statistic. We use χ^2 fitting to fit the data to the background-only hypothesis (H0) and the signal-plus-background hypothesis (H1). H0 has 1 degree of freedom (scaling of background) while H1 has 4 degrees of freedom (scaling of background, signal strength, signal width, higgs mass). We can then compute the p-value from the difference of the two reduced χ^2 values:

$$p = 1 - F_{\chi^2_{\Delta\nu}}(\Delta\chi^2) \tag{1}$$

where $F_{\chi^2_\nu}$ is the cumulative distribution function of the χ^2 distribution with ν degrees of freedom, $\Delta\chi^2 = \chi^2_{H0} - \chi^2_{H1}$ is the difference of the two reduced χ^2 values, and

$\Delta\nu$ is the difference in degrees of freedom between the two hypotheses.

The result is shown in Figure 5.

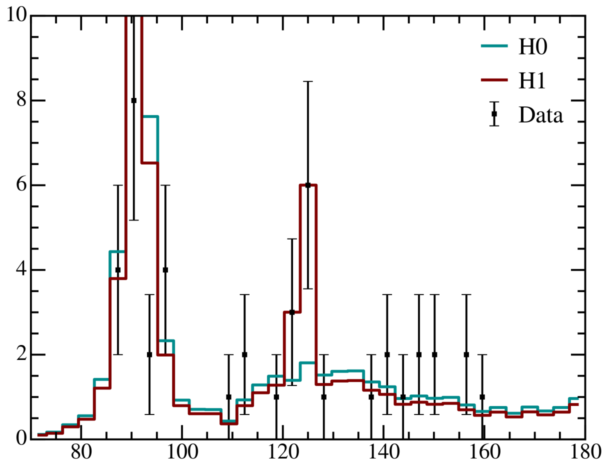


FIG. 5. Sample H0 and H1 likelihood distributions for $m_h = 125$ GeV. The blue line indicates the null hypothesis, and the red line indicates the alternative hypothesis.

The p-value is then used to calculate the significance of the excess, the number of standard deviations away from the null hypothesis, in signal according to the formula [10]:

$$Z = \Phi^{-1}(1-p) = \sqrt{2}\mathrm{erf}^{-1}(1-2p) \qquad (2)$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution, and erf^{-1} is the inverse error function. The significance is then used to determine whether the excess in signal is statistically significant.

We scanned mass range from 110 GeV to 160 GeV by approximating the signal as a Gaussian distribution on top of the background. We then performed statistical hypothesis testing using the profile likelihood ratio test to determine the significance of the excess in signal at each mass point. The results are shown in Figure 6, with highest significance of 2.33 σ at 125 GeV.

VI.1. Uncertainty Estimation

The statistical uncertainty comes from the finite size of the events recorded, which is estimated using the standard deviation of the number of events in each bin. The statistical uncertainty can be improved by measuring more events for longer duration with a higher combined luminosity of the dataset, which LHC did in 2015 and 2016 in Run 2 [1].

The major source of systematic uncertainty comes from the Monte Carlo simulation. In particular, the choice of parton distribution function (PDF) and the choice of the generator can affect the results. We used PDF4LHC

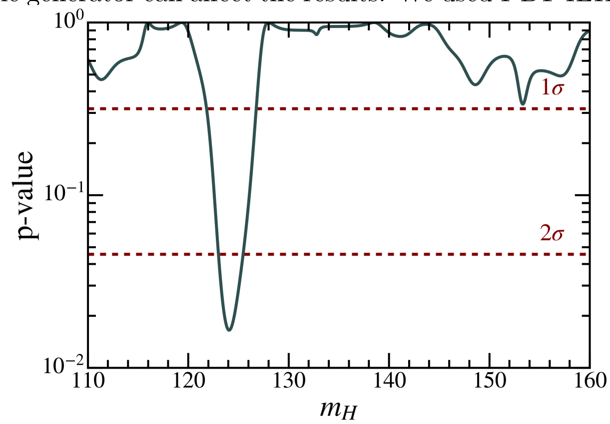


FIG. 6. Significance of the excess in signal as a function of mass. The red line indicates the expected significance, and the blue line indicates the observed significance.

working group recommendation [3, 6, 7, 15] for the choice of PDF.

VII. CONCLUSION

In conclusion, we have reproduced the analysis done in [8] and found an excess in signal of a Higgs-like particle around 125 GeV in the $H \to ZZ \to 4l$ search. The results are consistent with the Standard Model prediction of a Higgs-like particle with a mass of 125 GeV, with a significance of 2.33 σ. The discovery of Higgs boson completes the Standard Model of particle physics and confirms the Higgs mechanism, The results also provide a strong motivation for further studies of the Higgs boson and its properties, including its couplings to other particles, its role in electroweak symmetry breaking, and potentially more complicated Higgs sectors beyond the Standard Model.

ACKNOWLEDGMENTS

The author gratefully acknowledges their lab partner V. Tran for their invaluable assistance. The author also thanks the 8.13 teaching team for their guidance in the lab. This work was supported by the MIT Department of Physics.

[1] G. Aad et al. Combined Measurement of the Higgs Boson Mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ Decay Channels with the ATLAS Detector Using s=7, 8, and 13 TeV pp Collision Data. *Phys. Rev. Lett.*, 131(25):251802, 2023.

[2] S. Agostinelli et al. Geant4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[3] S. Alekhin et al. The PDF4LHC Working Group Interim Report. 2011.

[4] S. Alioli, P. Nason, C. Oleari, and E. Re. A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box. *JHEP*, 06:043, 2010.

[5] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D. L. Rainwater, and T. Stelzer. MadGraph/MadEvent v4: The New Web Generation. *JHEP*, 09:028, 2007.

[6] R. D. Ball and others (NNPDF Collaboration). Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology. *Nucl. Phys. B*, 849:296–363, 2011.

[7] M. Botje et al. The PDF4LHC Working Group Interim Recommendations. 2011.

[8] S. Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[9] C. Collaboration. Cms 2010 collision data, 2014.

[10] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[11] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[12] J. F. Gunion, H. E. Haber, G. L. Kane, and S. Dawson. *The Higgs Hunter's Guide*, volume 80 of *Frontiers in Physics*. Addison-Wesley, 1990.

[13] M. Hu. Higgs to 4l data analysis project, 2020. MIT Junior Lab Spring 2020.

[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] H.-L. Lai et al. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[16] MIT Department of Physics. Higgs to 4L Data Analysis Project. Technical report, Massachusetts Institute of Technology, Sep 2020. Dated: September 24, 2020.

[17] Particle Data Group. Status of Higgs Boson Physics. *Prog. Theor. Exp. Phys.*, 2022:083C01, 2022. 2023 update.

[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[19] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, Boulder, CO, 1995.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[21] T. Sjostrand, S. Mrenna, and P. Skands. Pythia 6.4 physics and manual. *JHEP*, 05:026, 2006.

[22] S. Weinberg. *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 6 2005.