

A 2.4σ observation of the Higgs Boson through the $H \rightarrow ZZ \rightarrow 4\ell$ golden channel

Vinh Tran

*Department of Physics and Kavli Institute for Astrophysics and Space Research,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA**

(Dated: April 20, 2025)

We present an observation of the Higgs boson utilizing data from the 2011-2012 run of the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). The analysis focuses on the $H \rightarrow ZZ \rightarrow 4\ell$ decay channel, which is considered the “golden” channel for Higgs boson detection due to its clean signature and relatively low background. We employ a combination of kinematic variables and machine learning techniques to suppress background events, resulting in a significant signal peak in the four-lepton invariant mass distribution. The analysis yields a 2.4σ significance for the Higgs boson signal, with a mass measurement of $m_H = 124.22 \pm 1.16$ GeV. This result is consistent with previous measurements and provides further evidence for the existence of the Higgs boson.

I. INTRODUCTION

By the beginning of the twenty-first century, most of the experimental evidence supporting the Standard Model, the quantum relativistic framework in which particles and phenomena are described as interactions of fundamental fields, had been established [e.g. 1–3]. The model has been successful in describing three of the four known fundamental forces in the universe: electromagnetism, the weak nuclear force, and the strong nuclear force, through the language of fundamental particles. These particles, which can be thought of as excitations of the underlying fields, include fermions (such as electrons and quarks) and force-carrying bosons (such as photons and gluons). One missing piece of the picture, however, was the Higgs boson, the particle whose associated field is responsible for imparting other particles with mass via the Higgs mechanism [4]. Such a particle has long been theorized [5], but not observed until 2012 [6].

Here, we present an observation of the Higgs boson using data from the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at CERN. We utilize the 2011-2012 datasets of proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 7$ TeV and 8 TeV [7]. The analysis focuses on the $H \rightarrow ZZ \rightarrow 4\ell$ (four-lepton) decay channel, which is considered the “golden” channel for Higgs boson detection due to its clean signature and relatively low background. Section II introduces the theoretical background of the Higgs boson decay into four leptons, as well as the background processes that can mimic this signal. Section III describes the CMS detector and the selection criteria for four-lepton events. Details on the Monte Carlo simulations used to model the signal and background processes are also provided. Section IV presents the details on the background suppression techniques employed in the analysis, including the use of kinematic variables and machine learning tech-

niques. Section V discusses the data analysis and results. Finally, we conclude with a discussion of the implications of our findings in Section VI.

II. FOUR-LEPTON DECAYS

II.1. $H \rightarrow ZZ$

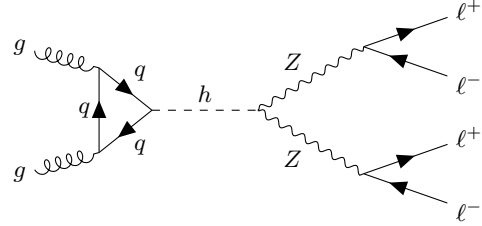


FIG. 1. The Feynman diagram of Higgs boson decay in the four-lepton “golden” channel.

The Higgs boson can be produced in proton-proton collisions through the gluon fusion process, in which two gluons from the colliding protons interact via a top or bottom quark loop to generate a Higgs boson. These Higgs bosons, with the expected mass of $m_H \simeq 125$ GeV, can then decay into two Z bosons, among other channels ($H \rightarrow WW$, $H \rightarrow bb$, and $H \rightarrow \tau\tau$) [8], one of which is considered off-shell, where the Z boson is produced with a mass below its nominal mass of 91.2 GeV. Each Z boson can then decay into a pair of lepton-antilepton, resulting in the final state of four leptons (electrons or muons). Figure 1 shows the Feynman diagram for this process. The decay of the Higgs boson into four leptons is regarded as the “golden” channel, as none of the final-state particles are neutrinos, which would result in missing energy and loss of information, or hadrons, which necessitate more complex modeling of jet fragmentation and hadronization. The invariant mass of the four-lepton system can be used to reconstruct the mass of the Higgs

* vinhtran@mit.edu

boson, allowing for a precise measurement of its properties.

II.2. Background events

Despite the clear signature of the Higgs four-lepton channel, its branching fraction, i.e. the fraction of decays into a particular channel, is only about $2.8 \pm 0.3\%$ [9]. On top of the low production rate of the Higgs boson, this results in other background processes, producing similar final states, dominating observations. Most of these events, such as the production of a Z boson through the Drell-Yan process [10] or lepton production of top-antitop interactions (in combination with jet misidentification, among other factors), are reducible through careful selection criteria for leptons, as discussed further in Section III and IV. However, the most significant background processes, involving the same ZZ decay channel as Higgs bosons, cannot be entirely eliminated. These processes, arising from interactions between quarks or gluons (as illustrated in Figure 2), necessitate further background suppression techniques, including analysis of Monte Carlo simulations and certain uses of neural networks.

III. EXPERIMENTAL DATA

The CMS detector consists of a central solenoid magnet, which generates a magnetic field of 3.8 T, and a series of detectors designed to measure the charge, momentum, and energy of particles [11]. From the center outwards, the detector includes a silicon tracker, which measures the trajectories of charged particles, an electromagnetic calorimeter (ECAL), where the energies of photons and electrons are measured, and a hadronic calorimeter (HCAL), which measures the energies of hadrons. The outermost layer is a muon system, which identifies and measures the momentum of muons. The detector geometry is described in terms of the transverse angle ϕ and the pseudorapidity η , which is defined as $\eta = -\ln(\tan(\theta/2))$, where θ is the polar angle of the particle with respect to the beam axis. For measurements of charged particles, pseudorapidity is limited to $|\eta| < 2.5$, while for neutral particles, it is limited to $|\eta| < 3.0$. For muons, the pseudorapidity is limited to $|\eta| < 2.4$.

The data utilized in this analysis were collected during the 2011-2012 LHC run, with a center-of-mass energy of $\sqrt{s} = 7$ TeV and 8 TeV, respectively. In order to guide the analysis and event selection criteria, we also utilize Monte Carlo (MC) simulations of relevant events to model the signal and background processes (details on such simulations can be found in [6]). We select only events with four leptons (electrons or muons) in the final state through enforcing the transverse and longitudinal

impact parameters with respect to the primary vertex to $|d_{xy}| < 0.5$ cm and $|d_z| < 1.0$ cm, respectively. These impact parameters are defined as the closest distance between the particle trajectory and the primary vertex in the respective plane of reference. The three-dimensional impact parameter significances of the leptons, i.e. the impact parameter signal-noise ratios, are enforced to be less than 4.0, which ensures that the lepton pairs produced from Z boson decays originate from the same primary vertex. We also require that the leptons are isolated from other particles in the events, which is achieved by requiring a maximum relative isolation of the lepton of 0.4 within the cone radius of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$. Here, relative isolation of the lepton is defined as the scalar sum of the transverse momenta of particles reconstructed within a distance ΔR of the lepton, divided by the transverse momentum of the lepton, with transverse momentum defined as $p_T = \sqrt{p_x^2 + p_y^2}$, where p_x and p_y are the transverse components of the momentum vector.

IV. BACKGROUND SUPPRESSION

Through the selection criteria detailed in Section III, we are able to select a sample of events with four leptons in the final state. However, as discussed in Section II.2, background processes that can mimic the signal remain as dominant sources of events. To suppress these background events, we utilize a combination of kinematic variables and machine learning techniques. The following selection criteria performance are evaluated within the context of MC simulations, serving as a guideline prior to the unblinding process.

First, we enforce charge conservation in the four-lepton system, which imposes the sum of charges in each event to be zero. This is a necessary condition for the production of a neutral Higgs boson. Additionally, leptons are required to have sufficient transverse momentum, specifically $p_T \geq 7$ GeV for electrons and $p_T \geq 5$ GeV for muons. As a result of these selection cuts, 90–95% of the Drell-Yan and top-antitop background events (depending on the exact decay channels) are removed, while the signal efficiency remains above 90%. Taking into account the relative process luminosities, these background contributions from these processes are reduced to a negligible level. Nevertheless, the irreducible ZZ background, of which only 20–30% of events are removed through the cut, remains the dominant source of background. As the remaining four-lepton events are mainly produced through the same ZZ decay channel, we pair up the particles into lepton-antilepton pairs and reconstruct the invariant masses of the Z bosons. The invariant mass is calculated as $m = \sqrt{(\sum_i E_i)^2 - (\sum_i \vec{p}_i)^2}$, where E_i and \vec{p}_i are the energy and momentum vector of the particle, respectively. Following [6], we require that the mass of

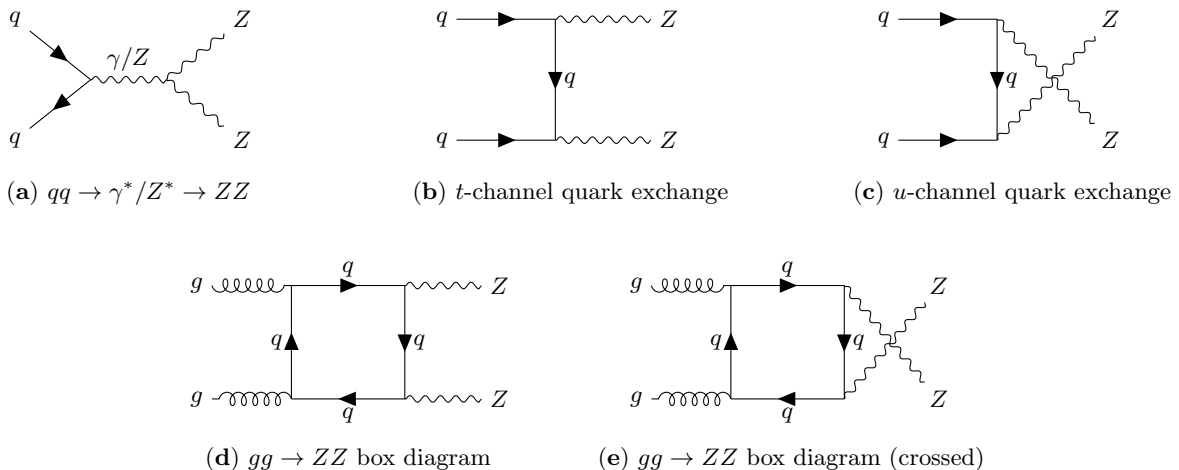


FIG. 2. Feynman diagrams of the non-reducible ZZ background processes via the $qq \rightarrow ZZ$ (a–c) and $gg \rightarrow ZZ$ (d–e) channels.

the heavier Z boson be within the range of 40–120 GeV, while the mass of the lighter Z boson is required to be within 12–120 GeV. This removes an additional 25% of the background events, while maintaining a signal efficiency of 95%.

In order to further remove background events, we utilize a fully-connected Graph Convolutional Network (GCN) [12] with each lepton as a node in the graph. The network is trained to perform event classification, distinguishing between signal and background. Each node is characterized by the particle type, encoded through a trainable embedding layer, and transverse momentum, which together serve as the input features for the model. The GCN architecture is designed to preserve symmetry between leptons while avoiding bias towards specific four-lepton invariant masses. Nevertheless, to avoid the 91.2 GeV Z boson peak, the network is trained exclusively on data within the invariant mass range of 100–160 GeV. We argue that this does not cause further bias in the selection, compared to the already-biased single choice of Higgs mass in the simulations. The optimal GCN model is selected based on the detection and false alarm rates, as well as the concentration of background removal. Using the model, we are able to remove an additional 70% of the background events, while maintaining a signal efficiency of 80%. The resulting prominence of the signal peak compared to the surrounding background count is around 5, which should be sufficient to observe the Higgs boson, given that the theory is correct.

V. DATA ANALYSIS & RESULTS

Figure 3 presents the distribution of the four-lepton invariant mass $m_{4\ell}$ for the selected events. The observed data from CMS appear to be consistent with the expected counts derived from the Monte Carlo (MC) sim-

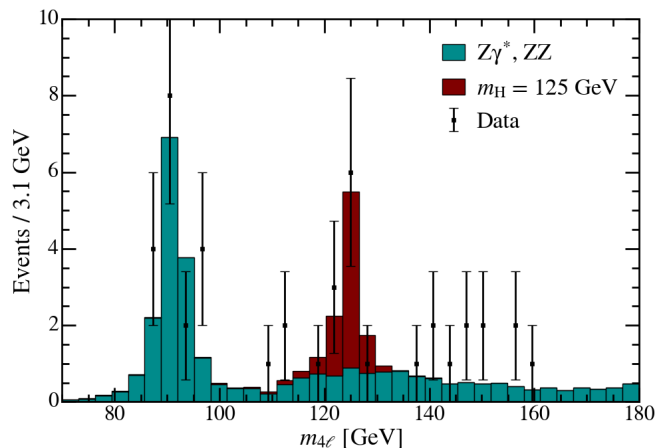


FIG. 3. Distribution of the four-lepton invariant mass $m_{4\ell}$ in the selected events. The red and cyan solid histograms represent the expected contributions from the Higgs boson signal with $m_H = 125$, GeV and the irreducible ZZ background, respectively. The observed event counts from the CMS 2011–2012 dataset are shown as black error bars.

ulated events. A bin width of 3.1, GeV is adopted for the histogram to balance resolution with statistical reliability, ensuring a reasonable number of events per bin. However, due to the overall low event counts, statistical uncertainties remain significant. To further investigate the potential Higgs boson signal, we conduct a hypothesis test. The null hypothesis corresponds to the absence of a Higgs boson signal, while the alternative hypothesis includes a Higgs signal component. Under the null hypothesis, the distribution of $m_{4\ell}$ is modeled using the background expectation from the MC simulations, scaled by a constant normalization factor. For the alternative hypothesis, the background model remains unchanged, and the signal is approximated by a Gaussian peak, centered at a variable Higgs mass m_H and characterized by

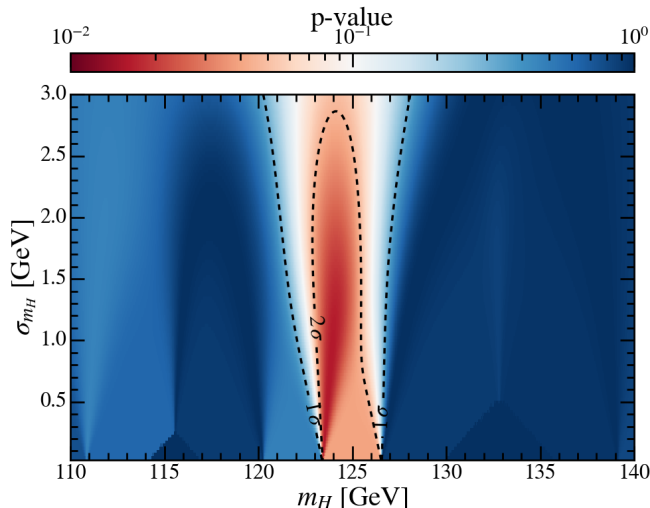


FIG. 4. The p-value as a function of the Higgs boson mass m_H and four-lepton signal characterized standard deviation σ_{m_H} . The dashed contours indicate the regions corresponding to 1σ and 2σ significance levels. Although σ_{m_H} is held fixed during each individual regression, the p-value is evaluated under the assumption that σ_{m_H} is a free parameter in the overall hypothesis test.

a standard deviation σ_{m_H} .

The p-value, representing the probability of observing a signal assuming the null hypothesis is true, is evaluated as a function of m_H . This involves fitting both signal and background models to the observed data. Due to the low event counts, we adopt a Poisson likelihood approach for the fitting procedure. To simplify the analysis, we perform a grid search over fixed values of σ_{m_H} within the range 0.1–3.1, GeV, as suggested by the observed distribution. The resulting p-values, shown in Figure 4, are computed from the difference in log-likelihoods between the null and alternative hypotheses. While σ_{m_H} is fixed during each fit, the p-values are interpreted under the assumption that it is a free parameter.

A clear minimum in the p-value distribution is observed at $m_H \simeq 123.8$, GeV, with a corresponding p-value of 0.016, indicating a local significance of approximately 2.4σ . A change in the standard deviation σ_{m_H} does not significantly affect the p-value, as the distribution remains relatively flat. This suggests that the observed signal is robust against variations in the width of the Gaussian peak. To obtain a more precise estimate of the Higgs mass and its associated uncertainty, we employ a bootstrapping procedure. The observed data are

resampled by perturbing the event counts according to their respective uncertainties, followed by a repetition of the analysis on each resampled dataset. This yields an estimated Higgs mass of $m_H = 124.22 \pm 1.16$, GeV, which is in good agreement with the expected value of $m_H \simeq 125$, GeV.

VI. DISCUSSION & CONCLUSION

In this study, we report an observation of the Higgs boson via the four-lepton decay channel, using data collected during the 2011–2012 run of the CMS experiment at the LHC. The analysis employs a combination of kinematic selection criteria and machine learning techniques to effectively suppress background contributions, leading to the emergence of a statistically significant signal peak in the four-lepton invariant mass distribution. A significance of 2.4σ is achieved, and the Higgs boson mass is measured to be $m_H = 124.22 \pm 1.16$, GeV. This measurement is consistent with prior results and constitutes additional evidence supporting the existence of the Higgs boson.

Despite the encouraging results, several limitations remain. First, the analysis is restricted to a single decay channel, which, while clean and well-reconstructed, may not capture the full phenomenological breadth of Higgs boson production and decay. Second, the modeling of background processes using Monte Carlo (MC) simulations shows some discrepancies with the observed data, as illustrated in Figure 3, suggesting potential shortcomings in the simulation or in the treatment of systematic uncertainties. Additionally, the focus on testing a single mass hypothesis may introduce bias, particularly in the presence of low statistics and imperfect background modeling. Future work should aim to refine the modeling of both signal and background processes, as well as expand the analysis to include additional Higgs decay channels and production modes. Furthermore, the application of more advanced machine learning methods may enhance the sensitivity of future searches by further improving background rejection and optimizing signal extraction.

ACKNOWLEDGMENTS

The author thanks his lab partner Y. Hu for collaboration in the experiment. Special thanks to the JLab teaching staff, P. Lugato, for their guidance and support, as well as CERN Open Data for providing the experimental data.

-
- [1] UA1 Collaboration, *Physics Letters B* **126**, 398 (1983).
 - [2] D0 Collaboration, *Phys. Rev. Lett.* **74**, 2632 (1995), [arXiv:hep-ex/9503003](https://arxiv.org/abs/hep-ex/9503003) [hep-ex].

- [3] DONUT Collaboration, *Physics Letters B* **504**, 218 (2001), [arXiv:hep-ex/0012035](https://arxiv.org/abs/hep-ex/0012035) [hep-ex].
- [4] S. Weinberg, *The Quantum theory of fields. Vol. 1: Found-*

- dations* (Cambridge University Press, 2005).
- [5] P. W. Higgs, *Phys. Rev. Lett.* **13**, 508 (1964).
 - [6] CMS Collaboration, *Physics Letters B* **716**, 30 (2012), [arXiv:1207.7235 \[hep-ex\]](#).
 - [7] N. Z. Jomhari, A. Geiser, and A. A. Bin Anuar, (2017), [10.7483/OPENDATA.CMS.JKB8.RR42](#).
 - [8] J. F. Gunion, H. E. Haber, G. L. Kane, and S. Dawson, *The Higgs Hunter's Guide*, Vol. 80 (2000).
 - [9] Particle Data Group Collaboration, *Phys. Rev. D* **110**, 030001 (2024).
 - [10] S. D. Drell and T.-M. Yan, *Phys. Rev. Lett.* **25**, 316 (1970).
 - [11] S. Chatrchyan *et al.* (CMS), *JINST* **3**, S08004 (2008).
 - [12] T. Kipf and M. Welling, [ArXiv abs/1609.02907](#) (2016).