Higgs to 4L Data Analysis Project

MIT Department of Physics

(Dated: September 24, 2020)

In this project you will use public data from the CMS experiment at CERN to demonstrate evidence of the Higgs boson. CMS detects particles created by collisions of very high energy protons. Properties such as the energy, momentum, and charge of these particles are recorded, and cuts can be made on these physical parameters to keep events that likely originated from a specificy. In this case, you will look for events that match the signature $H \to ZZ* \to 4l$. You will start with a dataset consisting of events that consist of 4 leptons being detected. You will then Monte Carlo predictions to determine cuts that will isolate events originating from a Higgs decay. After applying these cuts, you will then fit the Monte Carlo simulations to the data in order to subtract any remaining backgrounds, and demonstrate a detection of the Higgs boson.

I. INTRODUCTION

The Standard Model of particle physics is a mathematical model that describes the interactions of all fundamental particles (Fig. 1). It describes interactions between fundamental fermions, which consist of 6 quarks (including the up and down quarks that make up protons and neutrons), and 6 leptons (including three charged leptons known as the electron, muon, and tau, and three neutral leptons called neutrinos). The interactions are mediated by bosons, which are responsible for the four fundamental fundamental forces: the electromagnetic force (gamma), the weak force (W/Z), and the strong force (gluon). (There is a theorized fourth particle called the gravitino that would mediate the graviational force, but this is not included in the Standadr Model). The Standard Model has been extremely effective at making numerous and precise precise predictions of particle interactions.

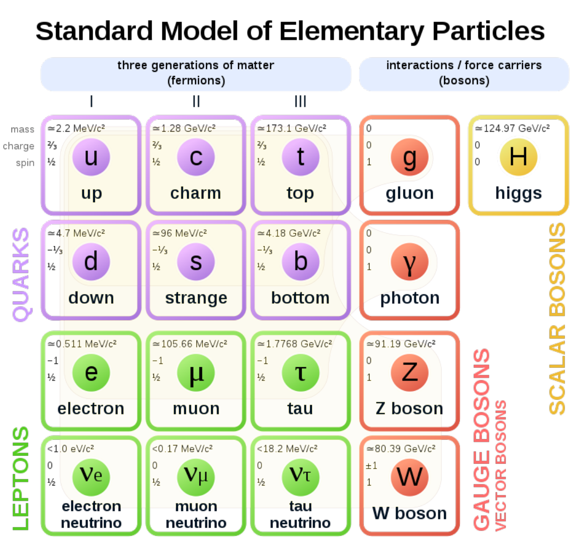

FIG. 1: Fundamental particles making up the Standard Model of Particle Physics [2].

However, there are stil missing gaps in the Standard Model. One such gap that was only recently filled is the origin of masses of the particles, such as the W and Z bosons. The equations describing the weak force were symmetric without masses, making it difficult to motivate why the particles would have masses. In the 1960's, a field was predicted by theorists that would create a mechanism for particles to have a mass. This was named the Higgs field, after the theorist Peter Higgs who helped predict it, and the corresponding particle is called the Higgs boson. This Higgs boson was detected by the CMS and ATLAS experiments at the LHC (Large Hadron Collider) in 2012.

The Higgs boson is produced at the LHC through the interactions of two gluons. It is then able to decay throgh a variety of channels. In this project, we focus on the "Golden Channel" of the Higgs decaying to two Z bosons, each of which then decays to two leptons. Charged leptons are much easier to detect than hadrons (particles made of quarks) because they deposit energy much more predictably in the detector. Once events with 4 leptons are selected, their energies and momenta can be used to reconstruct the kinematics of initial Higgs boson that decayed.

In this project you will be given the kinematic information for events that consist of four leptons. You will need to determine the most efficient cuts to isolate the Higgs to 4l signal and present a detection of the Higgs boson. The following sections will give a brief introduction to the CMS detector that was used to take the data, the general analysis procedures that will be followed, and details of how to access the data.

Note that this document is meant to serve as an overview of the project. See Miao Hu's ipython notebook from Junior Lab Spring 2020 on github for more details on the detector, analysis procedures, and code examples: `https://github.com/hmiaozh/MITJuniorLab/blob/master/HiggsTo4L/HiggsToZZTo4L.ipynb`.

II. THE CMS DETECTOR

The LHC is a particle collider that accelerates protons to energies of 6.5 TeV per beam, then collides them at four points around the ring where detectors are located.

CMS is one of these four detectors. CMS stands for Compact Muon Solenoid, named for the 4 T solenoid magnet around the interaction point that is used to bend particles in order to help determine their charge and momentum.

The detector consists of multiple layers, as shown in Fig. 3:

- Silicon Tracker: The Silicon tracker consists of silicon detectors that track the position of particles as they travel. This section is inside of the magnetic field from the solenoid, so particles are bent according to the ratio of their charge over momentum, q/p.

- Electromagnetic Calorimeter: Photons and electrons will be stopped by the electromagnetic calorimeter, and the deposited energy will be measured. Noting the bend of the particle in the silicon tracker helps determine the charge in order to identify the particles.

- Hadronic Calorimeter: Hadrons will be stopped here, and the energy will be recorded.

- Muon Chambers: Muons are much heavier than electrons, and will therefore travel through the electromagnetic calorimeter without being stopped. The Muon Chambers are used to track the curvature of the muon's path in a magnetic field in order to determine the muon momentum. The Silicon Tracker is unable to precisely measure this information because it is much smaller. The large mass of the muon leads to a low curvature that requires larger distances to measure precisely.

These outputs are enough to determine the particle type and kinematics for all particles emitted from each collision.

III. ANALYSIS PROCEDURES

Analysis of CMS data consists of several main steps. First, the tracks of particles through the detector are reconstructed, by using algorithms that cluster nearby hits together in order to determine the path through the tracker(s). This along with the energies measured in the calorimeters is sufficient to determine the particle type, as well as its kinematics. Multiple tracks that originate from the same "primary vertex" are then combined together into events.

Next, cuts are placed on the data in order to only keep events corresponding to a specific signal (i.e. the interaction of interest). For example, in our case, a cut is placed saying that the event consists of four charged leptons originating from the same primary vertex. Cuts can also be placed on the event kinematics, such as placing cuts on the pseudorapidigy (which corresponds to the angle of the emitted particle). These cuts are determined by using Monte Carlo simulations, where decays of particles and subsequent interactions are simulated computationally. Simulations are run for the signal and background processes. The expected distributions of kinematic variables can then be compared between the simulations for the signal and backgrounds, and cuts can be selected that are the most efficient at removing backgrounds while preserving the signal.

Note that in general it will not be possible to remove all background events. In particular, there are some "irreducible backgrounds" that are very similar to the signal and therefore can not be removed without also removing the signal. For example, there are other interactions which also produce two Z bosons. We would be unable to cut these since our cuts are designed to keep events resulting from the decay of two Z bosons.

Finally, the data remaining after cuts are applied are compared to the Monte Carlo simulations. This is the step where irreducible backgrounds can be dealt with. For example, in this case, the distribution of the four lepton invariant mass is calculated. The Higgs decay will give a peak around the Higgs mass: because it created the four leptons, conservation of momentum says that the sum of those four leptons will correspond to the Higgs mass. Background processes will have a different invariant mass distribution, and knowledge of that distribution from the Monte Carlo simulations can be used to subtract it in the region of the Higgs peak (see Fig. 2).

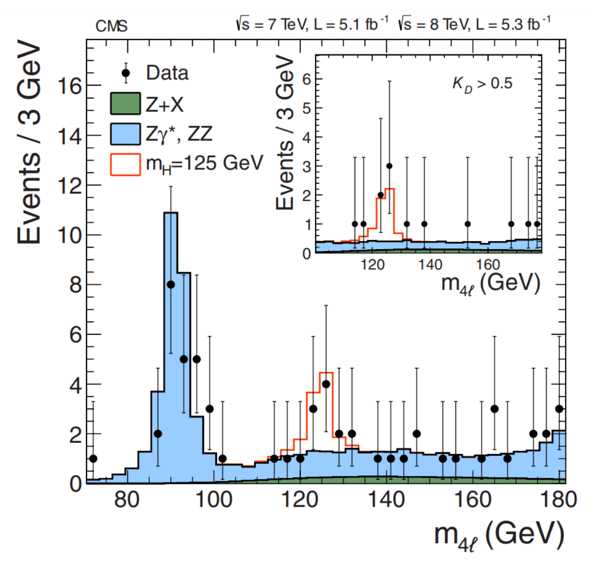

FIG. 2: Discovery plot for the Higgs boson. Background distributions from Monte Carlo are shown in green and blue. The red line indicates the excess corresponding to the Higgs peak [1].

In this project, the first step of getting event information has already been completed. In addition, preliminary cuts have already been made in order to give data consisting of four leptons (See Miao's github for more details). Your job is to determine further cuts that will work to filter out background events while keeping Higgs events. This can be done by considering characteristics of the interaction, such as properties of the intermediate Z boson decays. (You may want to read the discovery paper, especially section 5.2, to get some ideas: `https://arxiv.org/pdf/1207.7235.pdf`). You can test the efficiency of cuts by seeing how they effect the signal Monte Carlo and the background Monte Carlo. It also could be possible to teach a neural network to distinguish between signal and background events. Finally, after determining cuts, you will compare the Monte Carlo simulations to the resulting data in order to subtract irreducible backgrounds and measure the Higgs peak!

IV. DATA ACCESS

Data and Monte Carlo files for this project are accessible on github: `https://github.com/hmiaozh/MITJuniorLab/tree/master/HiggsTo4L`.

The data consists of csv files containing the run, event number, and the following information for each of the four leptons: Particle ID (PID), energy, momentum, pseudorapidity η, and azimuthal angle ϕ.

Sample scripts for working with the data are also provided in the ipython notebook on the github.

This project is based on the CERN Open Data Portal: `http://opendata.cern.ch/record/5500`.

V. RESOURCES

- Miao Hu's Junior Lab Spring 2020 github: `https://github.com/hmiaozh/MITJuniorLab/tree/master/HiggsTo4L`
- Original Higgs Discovery paper. Section 5.2 covers the $H4L$ channel: `arxiv.org/pdf/1207.7235.pdf`
- Info on Higgs decay channels: `http://www.scholarpedia.org/article/The_Higgs_Boson_discovery#Discovery_of_the_Higgs_boson`
- Brian Greene article about the Higgs discovery: `https://www.smithsonianmag.com/science-nature/how-the-higgs-boson-was-found-4723520/`

[1] S. Chatrchyanand et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30 – 61, 2012. URL: `http://www.sciencedirect.com/science/article/pii/S0370269312008581`, `doi:https://doi.org/10.1016/j.physletb.2012.08.021`.

[2] MissMJ. Standard model of particle physics. `https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`.

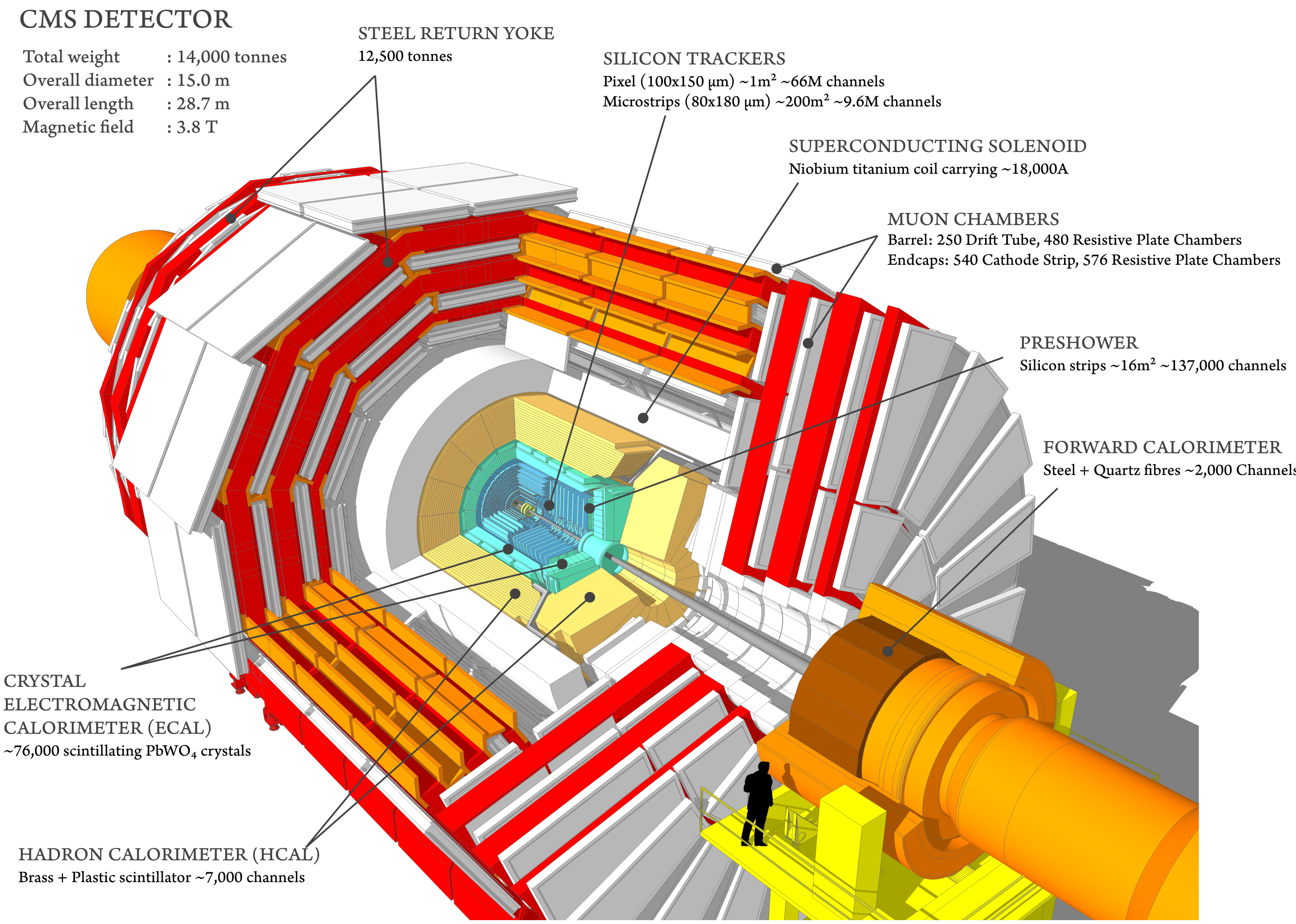

FIG. 3: Schematic of the CMS detector (`https://cms.cern/detector`).