**Discovering the Higgs Boson using Public CERN data: (report)**

1. Introduction

The Standard Model is a theory in physics that describes the fundamental particles and forces that make up the universe—except for gravity; namely fundamental particles and forces.

The importance of the Higgs boson lies in its connection to the Higgs field, which gives mass to other particles. Without the Higgs field, particles would be massless and unable to form atoms or matter as we know it. The discovery of the Higgs boson in 2012 confirmed this mechanism and filled a crucial missing piece in the Standard Model.

The interaction that the experiment was looking to measure was:

p + p → H → Z + Z\* → ℓ+ + ℓ- + ℓ′+ + ℓ′-

which indicates a proton collision where two gluons fuse into a Higgs boson. We then see the Higgs decay into a real and virtual (\*) Z boson, which goes on to decay into electrons or muons, plus their respective antiparticles. We can then detect and measure the energy of these.

The **Large Hadron Collider (LHC)** is the world’s largest particle accelerator, located at CERN. It smashes protons together at very high energies to explore fundamental physics. The **CMS (Compact Muon Solenoid)** experiment is one of the main detectors at the LHC. It played a crucial role in detecting the Higgs boson in 2012, providing strong evidence for the Higgs mechanism and confirming a key part of the Standard Model.

The goal of this piece was to analyse the public CMS data and search for the Higgs boson in the 4-lepton decay channel.

**2. Data and Methodology**

* A link to the data sets analysed can be found at this link to CERN’s page: [Higgs candidate events from CMS 2011 and 2012 open data release selected in the Higgs-to-four-lepton analysis example | CERN Open Data Portal](https://opendata.cern.ch/record/5200)
* The analysis of this involved combining this data. This was done via the Python Library Pandas - reading three CSV files into separate DataFrames, then combining all the rows into one single DataFrame.
* The invariant mass M of the 4-lepton system was used to determine the existence of the Higgs boson in the original experiment as well as here. The idea behind it is that the invariant mass of the 4 leptons totals to the mass of the exchange particle involved in its interaction.
* The Higgs bosons invariant mass was predicted to be 125GeV via theoretical work behind the standard model. Therefore, a clustering of total invariant mass around this mark would serve as a strong indicator to its existence. The histogram plotted aims to show this, plotting the total energy of the invariant system afterwards against the frequency of occurrence.

**3. Results**

* The histogram plot is below:

A graph of a number of blue and red lines

AI-generated content may be incorrect.

* The dashed red line has been marked at 125GeV, where we would expect to see a peak (via the standard model). This is evident by inspection – thus matching the theoretical predictions described about the Higgs boson.
* Other peaks and data points in this graph are due to the other exchanges that occurred; these largely involved other exchange particles such as the Z boson (91.2GeV).

**4. Discussion**

There are three main reasons why the ‘Higgs peak’ is not the tallest or sharpest peak as may be expect. The true natural width of the Higgs boson is 4MeV – but the resolution of the equipment used to measure leptons was larger than this, contributing to a broader peak. Direct Z boson pair production is also a by-reaction which produces a continuous distribution under and around the peak – so it appears less prominent. The reaction is also rare, so while you need a lot of reactions to display a peak, this also means a lot of irrelevant data points are collected.

**5. Conclusion**

In studying the Higgs boson through its decay into four leptons, I learned how physicists use the invariant mass of the 4-lepton system to identify the presence of the Higgs as a peak near 125 GeV. Although this peak is broad and not the tallest in the distribution, this is explained by factors such as detector resolution, background events, and limited statistics. The analysis revealed how interactions between fundamental fields—gluons, the Higgs field, Z bosons, and leptons—play out in high-energy collisions, with data providing the only window into these short-lived processes.

This highlights the crucial role of data analysis in particle physics, where detecting rare signals like the Higgs requires precise measurements, statistical modelling, and careful background subtraction. Moving forward, analysing larger datasets and performing detailed fits to the invariant mass distributions will improve the accuracy of Higgs property measurements and may even reveal signs of new physics beyond the Standard Model.

**6. Appendices (optional)**

The important snippets of the code used are below.

A screenshot of a computer

AI-generated content may be incorrect.

A white rectangular object with a white stripe

AI-generated content may be incorrect.