# LAB REPORT

## JOE BENTLEY

*School of Physics and Astronomy*
*Tutor: Kevin Ralley*

ABSTRACT. In this paper the higgs data is analysed and we find the higgs with this statistical significance etc.

## 1. Theory

### 1.1. Four Momentum.
The four momentum is similar to the classical three component momentum except generalised to four dimensional space-time, so it also has an energy component. In the data supplied the four-momentum is given in the form $(p_x, p_y, p_z, E)$, where $p_z$ is aligned along the beam axis.

### 1.2. Higgs Decay Channels.
The decay channel investigated in this experiment was the decay to a pair of photons through a top quark loop or W boson loop. This is significantly rarer than the other decay modes, but is used because the momenta of the photons can be measured much more accurately than other more massive particles, since the momenta of the photon is directly proportional to its energy, as the photon has no rest mass. (GET SOURCE)

### 1.3. Invariant Mass.
To find the mass of the particle(s) before a collision event, the invariant mass is used, which is the mass which remains unchanged through a Lorentz transform (GET SOURCE). The invariant mass of the products of an interaction are the same as the invariant mass of the input (???) particles, so by calculating the invariant masses of the resultant particles of the collision events, we can calculated a series of invariant mass values. Since the Higgs invariant mass is known (ROUGHLY?!?), the frequency of values in that window should be higher.

The invariant mass is calculated using the relativistic equation for energy, in natural units this is given by,

$$(1) \qquad m_0^2 = E^2 - |\mathbf{p}|^2$$

This is calculated in the parsing program by summing the four momenta from the event and then taking the scalar product of the resulting vector with itself. Minkowski space-time has signature $(+, -, -, -)$ so the scalar product has the effect of the above equation. We then take the square root of the scalar product to find the invariant mass.

## 2. PYTHIA

The GPL-licensed PYTHIA was used to generate the collision data used in this experiment.

[2] The event data is given in a flat text file format. The file consists of a series of events, which start with a header "Event $n$", the subsequent lines are the four momenta of the products of the collision event, then there is another line "Event $n+1$" with subsequent lines being four momenta events and so on.

## 3. Programming

The scripts used in this experiment were written in Python 3, and used the numpy (numeric Python) module for generating the histogram, and PyPlot (from matplotlib) for plotting the histogram and significance plots. Three different scripts are used in this experiment,

### 3.1. parse.py.
The script parse.py parses the event data in the format described in section 2 into individual event objects which each contain a list of four momenta objects. In these event classes we have methods which do useful operations like calculating the invariant mass of the event using the four momenta. The four momenta class has methods that allow us to easily calculate the transverse momentum, azimuthal angle, etc. The parsing scripts also contains the filters which are used to filter out Higgs events from background events, and are discussed in more depth in section 5. The invariant masses are calculated for each event in the event data file and then each invariant mass is written to a line in the output invariant masses file.

### 3.2. significance.py.
This scripts optimises the filters used on the Higgs and background data by using a series of parameters with the filters and testing them on the entire data set; the significance is taken for the parameters and this is used to find the best filter parameters, as discussed in section 6. This script parses and filters the data using the parsing module.

### 3.3. plot.py.
The plotting script generates the histogram of both the higgs and background event data together, called the combined mass histogram. The significance in the mass range (WHAT MASS RANGE) is calculated by counting the number of Higgs events in the mass range and the number of background events in the mass range and then using eq. 3 to calculate the

statistical significance. The script can also generate an invariant mass plot from the combined mass histogram.

## 4. WEIGHTING

The data given included $10,000$ signal events and $1,000,000$ background events. This is a much smaller ratio of signal to background events than would actually be observed in a real collider experiment, due to the low branching fraction of the $H \to \gamma\gamma$. The branching fraction is the fraction of all particles which decay through a given decay mode, in this case the fraction of $H \to \gamma\gamma$ decays over all Higgs decay modes. The theoretical ratio of the observed Higgs events to observed background events can be calculated by taking the ratio of the Higgs production cross section to the background cross section and multiplying it by the branching fraction of $H \to \gamma\gamma$,

$$(2) \qquad \frac{\text{num. H}}{\text{num. back}} = \frac{\sigma(\text{H prod}) \cdot B_f (H \to \gamma\gamma)}{\sigma(\text{back})}$$

where num. H is the number of signal events, num. back is the number of background events, $\sigma(\text{H prod})$ is the Higgs production cross section, $\sigma(\text{back})$ is the background cross section, and $B_f (H \to \gamma\gamma)$ is the branching fraction of Higgs to diphoton.

The branching fraction is given by $B_f (H \to \gamma\gamma) = 0.00225$, and the cross sections are given by $\sigma(\text{H prod}) = 17.35\,\text{pb}$, and $\sigma(\text{back}) = 140\,\text{pb}$. [1] (JUSTIFICATION NEEDED).

## 5. FILTERS

There are many more invariant masses from the background events than there are from the signal (Higgs) events. In fact after the weighting, there are roughly 100 Higgs events to $100,000,000$ background events, so in the combined invariant mass histogram the Higgs is practically invisible. Thus filtering needs to be applied for there to be any chance to be able to observe the Higgs in the invariant mass histogram. This is implemented in the form of event selection to remove events which are clearly not Higgs events. There are a few kinds of filtering EXPLAIN THEM.

5.1. **Transverse Momenta Filtering.** The transverse momenta is the component of the momentum perpendicular to the beam axis. Since the $z$ component of the four-momentum, $p_z$ is aligned along the beam axis, the transverse momentum is given by $p_T = \sqrt{p_x{}^2 + p_y{}^2}$. This is filtered in two passes over both the signal and higgs events. Given an event with four momenta, first the event is discarded unless it has at least one four momentum above a given threshold, then all four momenta in the event below a certain threshold are discarded.
(JUSTIFICATION)

5.2. **Energy.** (Filter the energy, not sure if we will see results)

5.3. **Angular.** (Filter the difference in azimuthal angle and pseudorapidity squared)

5.4. **Final Selection.** Finally all but the two highest transverse momenta photons are discarded, and the two remaining are used to calculated the invariant mass.
MENTION AZIMUTHAL SELECTION

## 6. OPTIMISATION OF FILTERS

To find the filter parameters that give the best chance of finding the Higgs we used the formula for the statistical significance of the higgs,

$$(3) \qquad \Sigma = \frac{\text{no. signal}}{\sqrt{\text{no. signal} + \text{no. background}}}$$

Each filter is applied to the signal (Higgs) and background events and then the above equation is used to find the statistical significance of the Higgs in those events by counting the number of signal and background events left after filtering. To find the best parameters to use for the filters a range of different parameters are tested for each filter. For example for the transverse momenta filter, the threshold is originally applied from $0\,\text{GeV}$ to $700\,\text{GeV}$ in steps of $100\,\text{GeV}$. The value that gives the best statistical significance is then chosen as the current candidate for the best optimised value. Then the value is further optimised by taking the currently best optimised value and then calculating the statistical significance for the values around it, for example checking the statistical significance at $\pm 10\,\text{GeV}$ the current value for the above example. After

repeating this process a few times (HOW MANY TIMES?) the most optimum filter values to see the most Higgs event in the invariant mass histogram are found.

INCLUDE SIGNIFICANCE PLOT!!!

## 7. Invariant Mass Histogram

The numpy library is used to generate an combined (signal and background) invariant mass histogram from the invariant masses calculated in the parsing script. The statistical significance of the signal (Higgs) is calculated using eq. 3 taken over a certain mass window (WHAT MASS WINDOW?!?), and gives a measure of how sure we can be that it is the Higgs in this mass range.
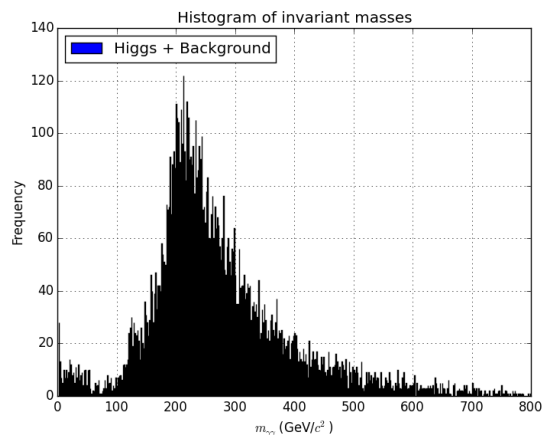


Figure 1. Our awesome invariant mass plot yo!

## References

[1] S. Dittmaier et al. Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *arXiv*, 2011.

[2] Torbjörn Sjöstrand; Stefan Ask; Jesper R. Christiansen; Richard Corke; Nishita Desai; Philip Ilten; Stephen Mrenna; Stefan Prestel; Christine O. Rasmussen; Peter Z. Skands. An introduction to pythia 8.2. `http://home.thep.lu.se/~torbjorn/Pythia.html`.