Precision Measurements of $H \rightarrow b\bar{b}$ Coupling from e^+e^- Collisions

Laura Lai

August 14, 2015

Abstract

The proposed 80 to 100 km Future Circular Collider will be an ideal setup for studying properties of the newly-discovered Higgs boson with much more precision. Physics of e^+e^- collisions are accessible at a center-of-mass energy of 90 to 350 GeV and high luminosity. The lepton collider, FCC-ee, may be used as an intermediate stage before pp collision to study decay channels of the Higgs at a center-of-mass energy of 240 GeV. The objective of this project is to search for the uncertainty of $H \rightarrow b\bar{b}$ coupling via event generation and simulation. Pythia8 was utilized in event generation, Delphes for CMS detector simulation, and ROOT for data analysis. After improving the b-tagging efficiency formula in CMS detector simulation, a C++/ROOT analysis was performed on the simulated data. Through this method, the $H \rightarrow b\bar{b}$ coupling was measured to be 6% with an integrated luminosity of $500 fb^{-1}$ with $Z \rightarrow \mu^+\mu^-$, $H \rightarrow b\bar{b}$ decay.

1 Introduction

1.1 Future Circular Collider (FCC) and Motivation

With the discovery of the Higgs boson, the Standard Model is complete. However, Higgs measurements must be made more precisely to be sensitive to new physics. The objective of the Future Circular Collider (FCC) is to explore new high-energy frontiers after the completion of the Large Hadron Collider (LHC) experimental runs. The FCC study mainly emphasizes the research and development of a hadron collider in an 80 to 100 km ring that is intended to reach a center-of-mass energy in the order of 100 TeV. There is an interest in exploring the possibilities of new physics even after the high-luminosity upgrade of the LHC, as well as exotic decays and precision measurements of the Higgs boson, other particles, and the properties of strong interaction.

There is a strong physics argument for colliding leptons in the FCC. The FCC-ee (an electron-positron collider) will reach center-of-mass energies of 90 to 350 GeV and will serve as an intermediate step in the project. To maximize the Higgs boson production cross-section, a center-of-mass energy of 240 GeV is selected. At this energy, the production of the Higgs boson with the Z boson becomes favorable. In order to study the Higgs with high precision, we need to reduce the systematic uncertainties that arise from hadron colliders. As measurements of Higgs properties is expected to have a precision of within a few percent, an e^+e^- collider would be an excellent candidate for such studies.

This project studies the behavior of a Compact Muon Solenoid (CMS)-like detector in an e^+e^- collider such as the FCC-ee. CMS is known for its performance in proton-proton and heavy ion collisions and is able to reconstruct and identify individual stable particles through particle-flow event reconstruction at excellent energy resolutions. Because of this, CMS can serve as an initial choice for prospective studies of FCC.

2 Methods

The project has three components:

- Deriving and implementing a new b-jet detection efficiency formula in Delphes simulation.
- Reconstruction of Z and H invariant and missing masses in C++/ROOT analysis.
- Fitting the data using a RooFit analysis to determine the uncertainty of $H \to b\bar{b}$ coupling.

Acquaintance with Pythia8, Delphes, and ROOT were required to complete the study. Pythia8 was used to generate high-energy physics events, in this case the results of e^+e^- collisions, and output the results into HepMC format, which are commonly used in HEP studies. The HepMC files were then used in Delphes, a software that simulates the response of a multipurpose detector in a parametrized fashion, using a Delphes detector card with CMS settings. Finally, C++/ROOT macros were written to evaluate b-jet detection efficiency, the number of events produced, the mean number of events, and statistics of the interaction.

2.1 b-Jet Detection Efficiency

In high-energy particle physics, b-tagging is an algorithm to identify jets originating from bottom quarks.

In the CMS Delphes card, the settings of the detector are programmed such that the simulation behaves as the CMS detector would in experimental runs. We updated the formula to reflect our belief that a looser working point could better serve our analysis. We selected a "medium" b-tag working point for the CSV algorithm due to its high efficiency and relatively low mistag rate. By extrapolating the data points from b-jet identification studies using the CSVM algorithm, a simple C++/ROOT analysis found the roots of the polynomial representation of the plot in Figure 2. The new b-jet efficiency formula was found to be:

$$\begin{aligned} \epsilon_{b-tag} = 6.08678 \times 10^{-1} + (1.70162 \times 10^{-3})p_T - \\ (5.72401 \times 10^{-6})p_T + \\ (4.75956 \times 10^{-9})p_T \end{aligned} \tag{1}$$

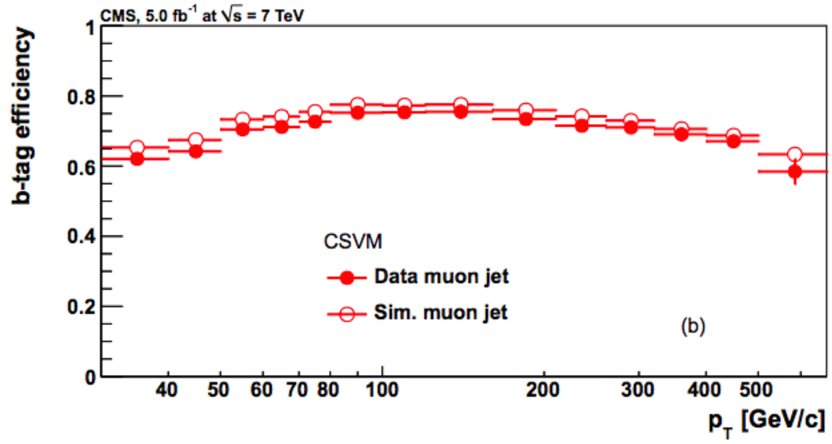


Figure 1: This plot was used to extrapolate the medium working point b-jet detection formula. The results include the efficiencies of b-jet detection using the CSVM tagger.[3]

Equation 1 was implemented in the Delphes CMS detector card, where the b-tagging acceptance was changed to $|\eta| < 2.4$ and $p_T > 15.0 GeV/c$.

Figure 2a shows the results of analysis with the old efficiency formula, and Figure 2b with the new. There was an immediate increase in the number of b-jets tagged after implementing the new formula. With the default efficiency formula, the majority of the events detected were events with one b-jet tagged. Many more events with two b-tagged jets were detected after using the new formula. In the analysis, the distribution of all jet events were taken into account, and this was compared to the number of b-jets per event. By using the selection criteria of $p_T > 20.0 GeV/c$ and $|\eta| < 2.4$ in the C++/ROOT analysis, it is confirmed that all instances of b-jets that meet this criteria in simulation are tagged.

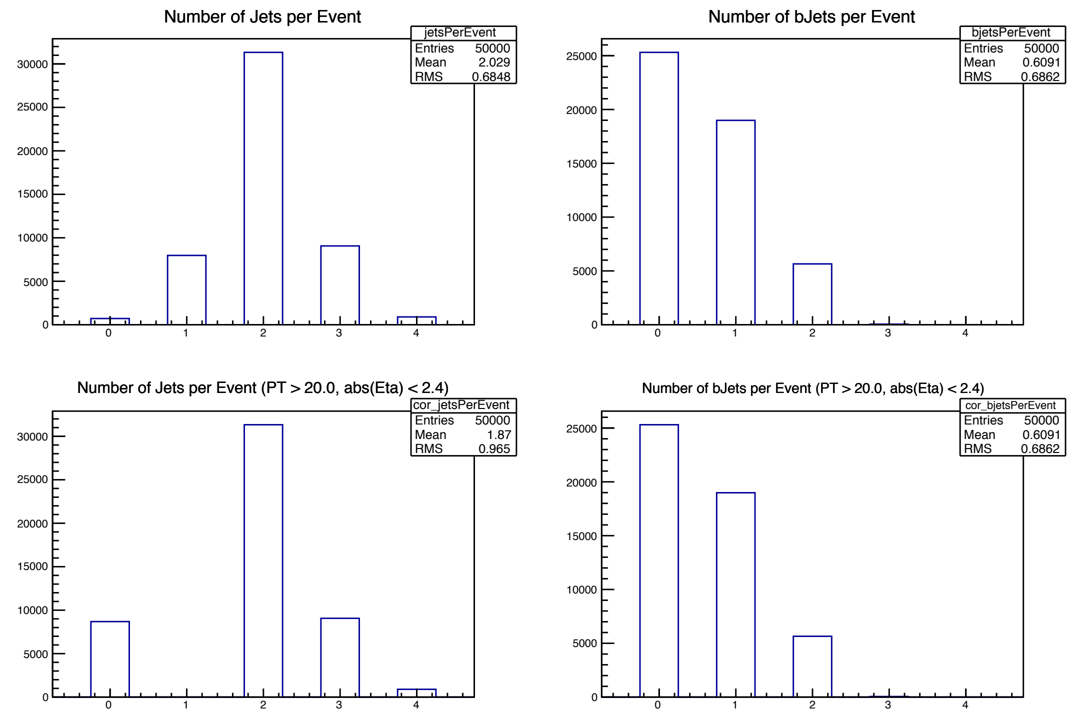
2.2 Event Generation and Simulation

To generate the desired events, Pythia cards were built so the processes in Figure 3 could be studied. For both the ZZ background and ZH signal, 100,000 e^+e^- collisions at $\sqrt{s} = 240 GeV$ were generated.

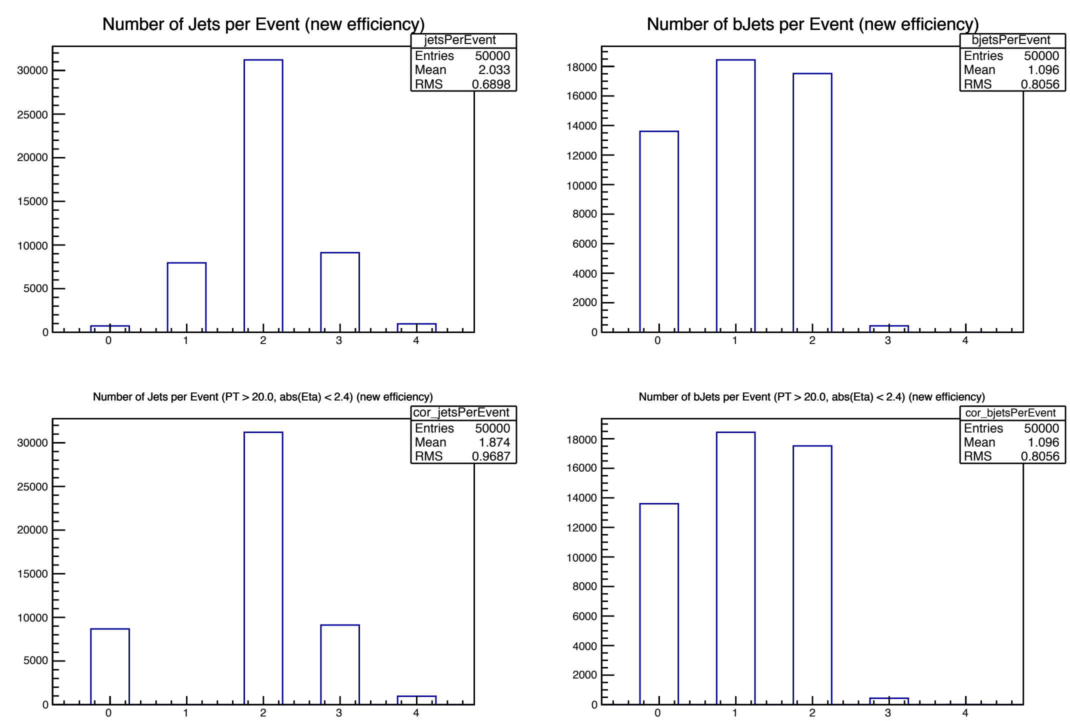
In the generation process, the cross-sections of the signal and background are provided, which are used in the scaling of the mass reconstruction plots to our desired integrated luminosity of $500 fb^{-1}$. The cross-sections of the two processes from Pythia8 simulation were found to be:

$$\begin{aligned} \sigma_{ZH} &= 3.962 \times 10^{-12} mb \\ \sigma_{ZZ} &= 4.753 \times 10^{-11} mb \end{aligned} \tag{2}$$

The HepMC files for ZZ and ZH events were processed in Delphes using the CMS card with the new b-tag efficiency formula. After simulating the behavior of the events in a CMS-like detector, the results are output in the form of ROOT trees that may later be analyzed by macros.

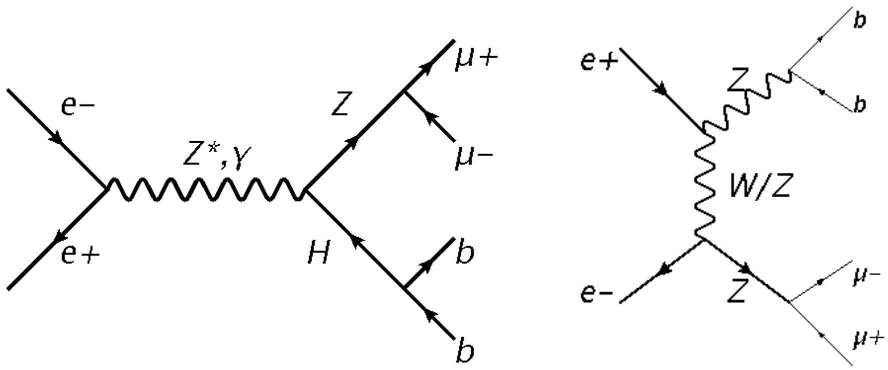


(a) Old b-jet detection efficiency formula.



(b) Results with new b-jet detection efficiency.

Figure 2: Histograms highlighting the differences between the old and new efficiency formulas.



(a) Signal (b) Background

Figure 3: Processes studied in experiment, simulated from e^+e^- collisions in Pythia8.

2.3 Reconstruction of Mass Events

Two algorithms were used to reconstruct the $H \to b\bar{b}$ mass events: invariant mass and missing mass. The invariant mass formula used was:

$$M_{inv}^2 = E^2 - \|\mathbf{p}^2\| \tag{3}$$

The missing mass formula used was:

$$M_{miss}^2 = (s - E_{\mu_1+\mu_2})^2 - (\overrightarrow{p}_{\boldsymbol{\mu_1}} + \overrightarrow{p}_{\boldsymbol{\mu_2}})^2 \tag{4}$$

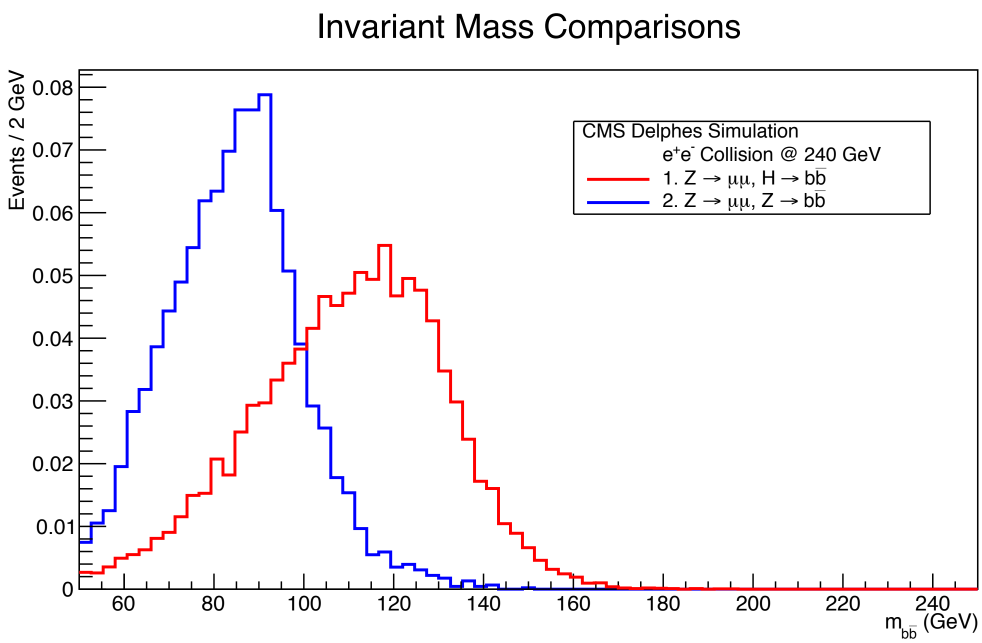
Events with at least two muons were selected with a $p_T > 15.0 GeV/c$ and a range for m_z at 60 to 120 GeV. A further cut was made such that these events must also have two b-tagged jets, requiring that $p_T > 15.0 GeV/c$ and the absolute value of pseudorapidity is less than 2.4, using the CSVM algorithm. The results of both reconstructions are found in Figure 4.

It is clear that the separation between the background and signal is more well-defined in the missing mass plot than that of the invariant mass. The missing mass has higher resolution spectra due to the use of a known value of center-of-mass energy. Both plots were scaled to a desired luminosity of $500 fb^{-1}$ (5), and the scaled plot in Figure 5b was used in a statistical fit.

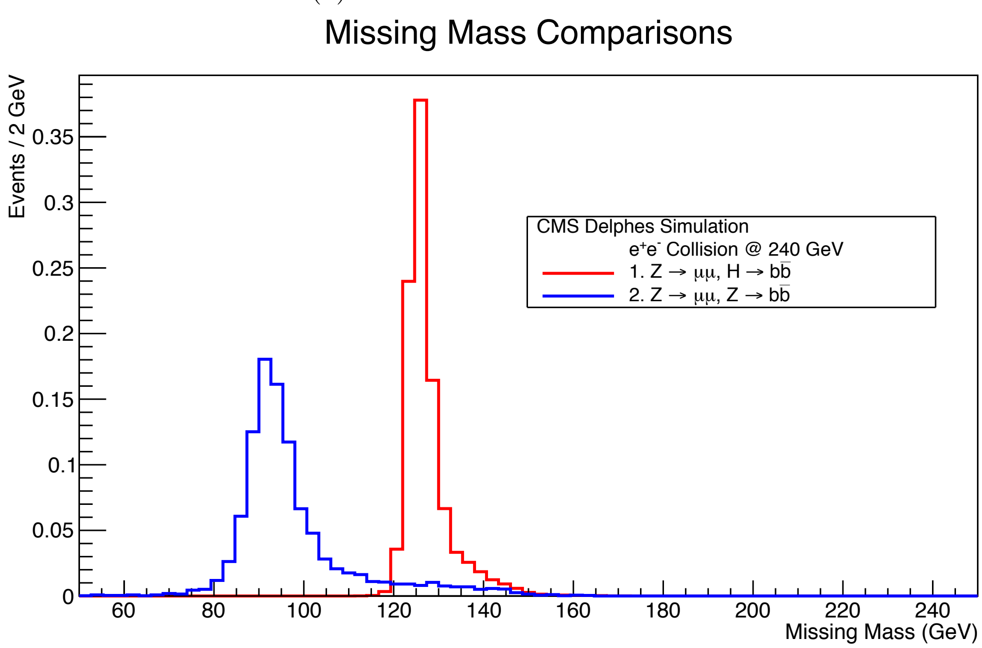
2.4 Fit and Statistical Evaluation

Figure 8 summarizes the results from the fit. Once the mass events were selected and reconstructed, a fit was made over the simulated data to ultimately evaluate the uncertainty of $H \to b\bar{b}$ coupling. The sum of both the ZH signal and ZZ background histograms, found in Figure 7, was smeared using RooFit and used as a toy Monte Carlo data set in the RooStats analysis. The signal is fit to a Crystal Ball function and the background to a first order polynomial. The following fit parameters provided an optimized result:

- Range: 110-150 GeV

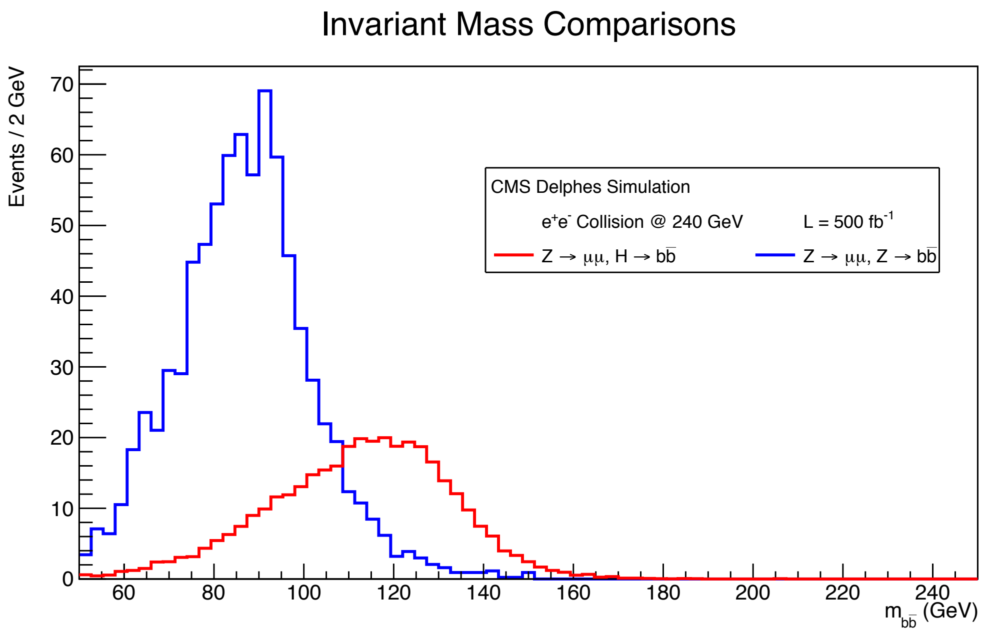


(a) Invariant Mass Reconstruction

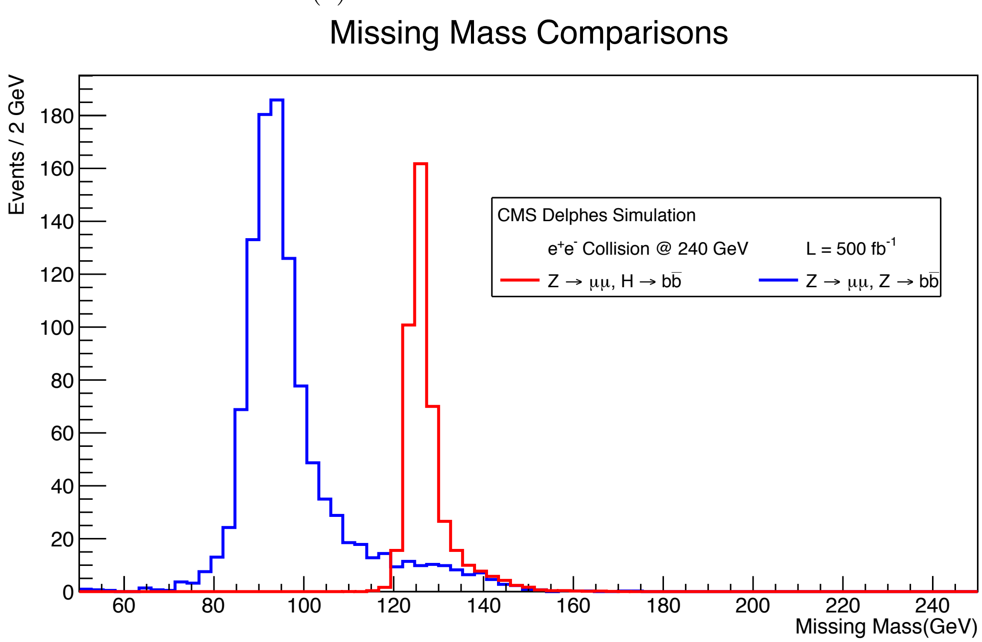


(b) Missing Mass Reconstruction

Figure 4: Results of mass reconstruction via ROOT macro.



(a) Invariant Mass Reconstruction



(b) Missing Mass Reconstruction

Figure 5: Results of mass reconstruction via ROOT macro, scaled to $L_{int} = 500 fb^{-1}$.

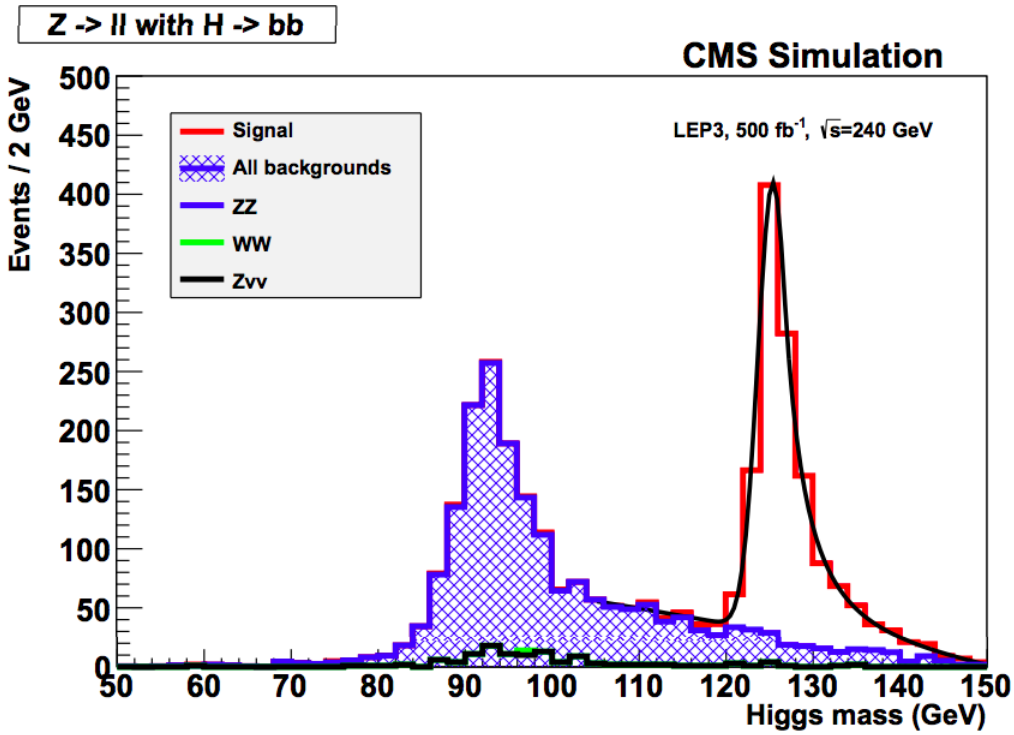


Figure 6: Plot of b-tagging efficiency as a function of p_T as a result of Combined Secondary Vertex Medium (CSVM) tagging algorithm from a study on the identification of b-jets [3]

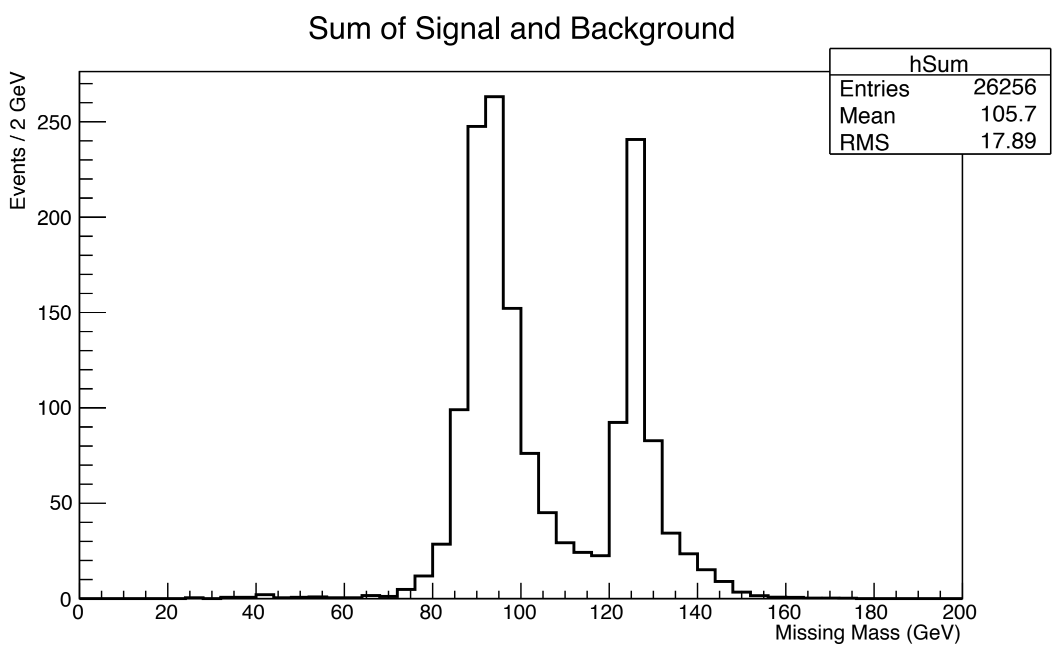


Figure 7: The sum of the missing mass signal and background was smeared and used as a toy MC to create a fit

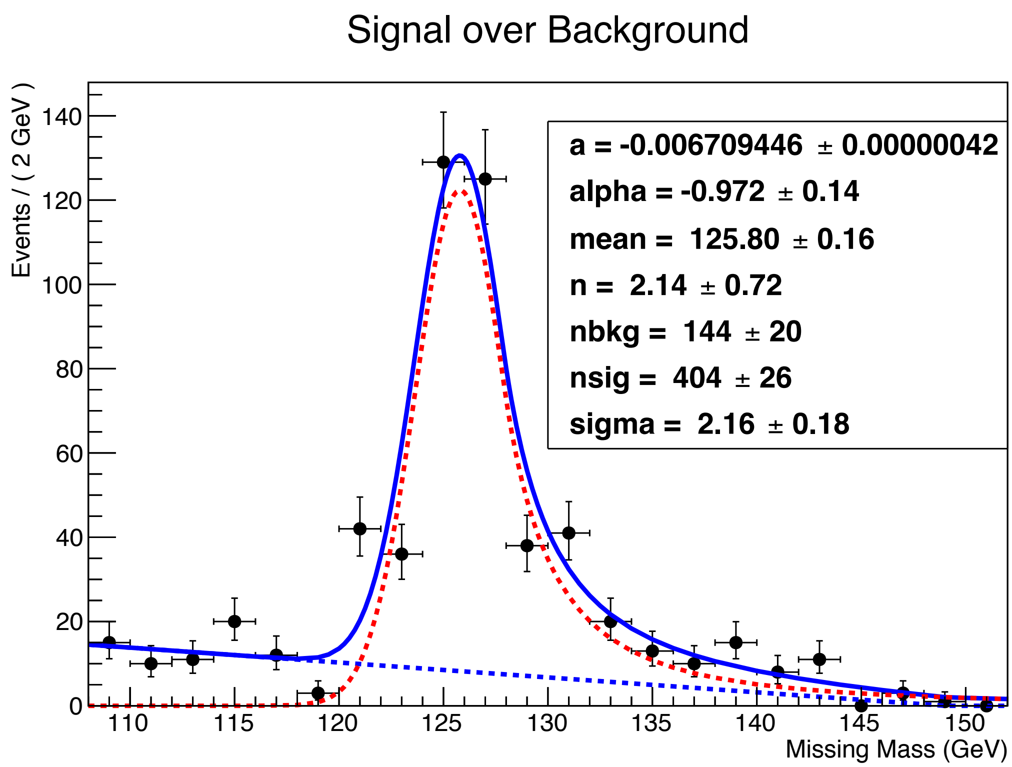


Figure 8: Results from RooStats/RooFit analysis.

- Signal: Crystal Ball
  - $N_{exp} = 425$
- Background: Polynomial
  - $N_{exp} = 144$

The expected number of events for the signal and background were found by integrating over the 110-150 GeV range for both scaled signal and background histograms. The results shown in Figure 8 show that the number of background and signal events from the fit are consistent with integration. From this fit, the precision of the $H \to b\bar{b}$ coupling is found to be 6%.

3 Conclusion

The precision of $H \to b\bar{b}$ coupling was found to be 6% using these methods. The results from "Prospective Studies for LEP3 with the CMS Detector" yielded a precision of 3.1% after fitting the signal to a Crystal Ball function and the background to a third order polynomial. A multivariable analysis might improve the precision to 2.7% by reducing the background by an order of two[1]. If this analysis had included $Z \to e^+e^-$ background, 4% precision is expected, which is very close to the value found in the full simulation study. If we were to scale the plots to $L_{int} = 3ab^{-1}$, the proposed luminosity of FCC-ee, the precision would scale down to 1.7%, from $Z \to \ell\ell$ $H \to b\bar{b}$ data only. These are two points that are possible, and ideal, to explore for further analysis in this project.

To improve this study, performing statistical analysis over 10^6 or greater number of events may improve the value of the precision. The fitting method could also benefit from an improved Crystal Ball fit over the toy MC as well as taking a look at a higher order polynomial fit for the background. Future studies may consider the behavior of other multipurpose detectors in the FCC-ee setup, such as an ILD detector more suitable for e^+e^- collisions, and determine which aspects of these detectors are beneficial for these precision measurements.

References

[1] Azzi, P., Bernet, C., Botta, C., Gomez-Ceballos, G., Janot, P., Klute, M., Lenzi, P., Malgeri, L., Zanetti, M., "Prospective Studies for LEP3 with the CMS Detector," *The Compact Muon Solenoid Experiment: CMS Note*, October 2012.

[2] CERN Future Circular Collider Collaboration, *Future Circular Collider Study*, https://fcc.web.cern.ch/Pages/default.aspx, 2014.

[3] The CMS Collaboration., "Identification of b-quark jets with the CMS experiment," *The Journal of Instrumentation*, May 2013.

[4] Pythia8 Documentation. http://home.thep.lu.se/~torbjorn/pythia81html/Welcome.html

[5] Delphes Documentation. https://cp3.irmp.ucl.ac.be/projects/delphes

[6] RooStats/RooFit Documentation. https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome