ZH Jet Pairing Methods for Higgs Self-coupling Sensitivity Optimization at a Future Higgs Factory

Phillip Ionkov
Department of Physics, Columbia University, New York, NY 10027

Under the direction of

Abraham Tishelman-Charny
Department of Physics, Brookhaven National Laboratory, Upton, NY 11973

August 9, 2023

Abstract

In 2012, the ATLAS and CMS collaborations, based at the Large Hadron Collider (LHC) at CERN, experimentally confirmed the existence of a particle consistent with the Standard Model (SM) Higgs boson, the final missing piece of the SM. The Higgs boson arises from excitations in the Higgs field, which gives mass to other SM particles. The next step in verifying the SM is to perform precise measurements of the Higgs boson's parameters, including the Higgs self-coupling, which has a direct impact on the shape of the Higgs potential. Verifying that the Higgs potential is SM-like or not would provide insight into the stability of the vacuum, but measuring the Higgs self-coupling is difficult because the physical processes showing observable changes due to its value are very rare. A proposed post-LHC particle collider, the Future Circular Collider (FCC), offers improved sensitivity to the Higgs self-coupling. This depends on many measurements of the Higgs boson's properties, but this project focuses on an unexplored decay channel of the associated production of a Z boson and a Higgs boson (ZH): the case in which both the Higgs and the Z bosons decay into quarks, which is in fact the most common final state for the ZH signal. This project, conducted in the Physics department at Brookhaven National Laboratory, investigates multiple methods toward identifying Z boson candidates by analyzing simulated FCC data using Z decay product masses and jet identification scores as metrics. Then, the recoil mass is calculated and used as a metric for evaluating sensitivity toward the Higgs self-coupling. These studies further the DOE's mission of fundamental research as they contribute toward the physics case for the FCC project. Through this project, I learned development in C++ and Python, and how high energy physics data is analyzed and presented.

1 Introduction

The Standard Model (SM) is currently the best explanation that particle physicists have for the fundamental workings of the universe. Experimental studies have generally agreed with predictions made by theorists under the laws of the SM, so it continues to be the basis of physics predictions. In 2012, the ATLAS and CMS collaborations, based at the Large Hadron Collider (LHC) at CERN, experimentally confirmed the final missing piece of the SM by observing a particle consistent with the SM Higgs boson [1, 2].

The Higgs boson arises from excitations in the Higgs field — massive particles obtain their mass through interactions with the Higgs field. This mechanism comes from a spontaneous breaking in the symmetry of the Higgs potential. Figure 1a depicts the shape of the Higgs potential. When a Higgs particle goes to a lower energy state, symmetry is broken, and the particle exists in a minimum of potential energy. It is unclear whether this minimum is local or global. The SM predicts that this minimum is local, and thus the universe is in a metastable vacuum state (Figure 1b). This prediction greatly depends on the shape of the Higgs potential, which is yet to receive a sensitive measurement. Studying the Higgs potential ultimately helps build an understanding of spontaneous symmetry breaking in general.

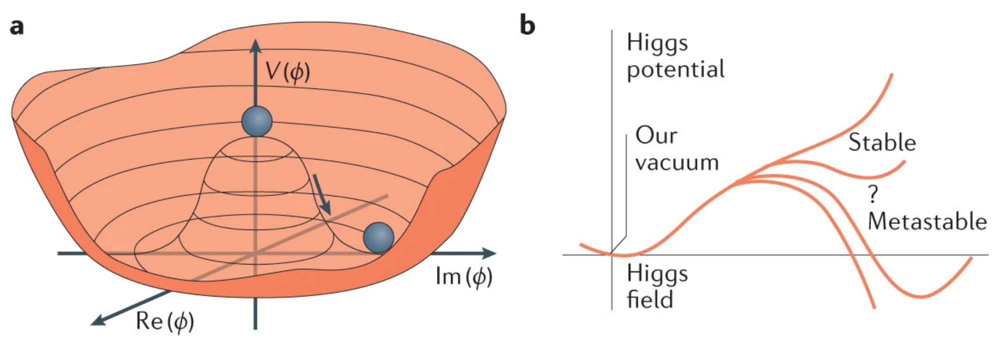

Figure 1: a) Higgs potential [3] b) Vacuum stability possibilities [4]

The strength of the Higgs self-coupling can be evaluated via Higgs pair production, the production of two Higgs bosons in a single process. However, Higgs pair production is a rare process, and current collider experiments are not sensitive enough — the necessary precision is $\sim 5 - 10\%$ [5], where lower is better, and the best precision that will be achieved at the upgraded High Luminosity Large Hadron Collider (HL-LHC) will only be 50% [6]. Therefore, one physics goal of a future collider is a sensitive analysis to the Higgs self-coupling.

1.1 The Future Circular Collider

The Future Circular Collider (FCC) is a proposed next-generation particle accelerator at CERN [7]. FCC plans to be the highest luminosity collider ever made, allowing for extremely sensitive measurements. In its initial stage, it will work as an electron-positron collider (FCC-ee), operating over increasing center-of-mass energies ($\sqrt{s}$) over time [8]:

- $\sqrt{s} \approx 90$ GeV to study Z boson physics
- $\sqrt{s} \approx 160$ GeV to study W boson physics through WW production

⇒ $\sqrt{s} \approx 240$ **GeV to study Higgs physics through the associated production of a Z boson and a Higgs boson**

- $\sqrt{s} \approx 350$ GeV to study top quark physics through $t\bar{t}$ production, as well as continuing to study Higgs physics

After its electron-positron phase, FCC will be upgraded to a hadron collider: "FCC-hh". FCC-hh plans to operate at ∼100 TeV, an enormous increase in energy compared to other accelerators, such as the LHC, which currently operates at 13.6 TeV. However, the technology necessary for FCC-hh is far from being close to usable.

This project explores the Higgs physics stage, or “Higgs factory” stage, of FCC-ee as a candidate for conducting a sensitive measurement of the Higgs self-coupling.//

1.2 The ZH process

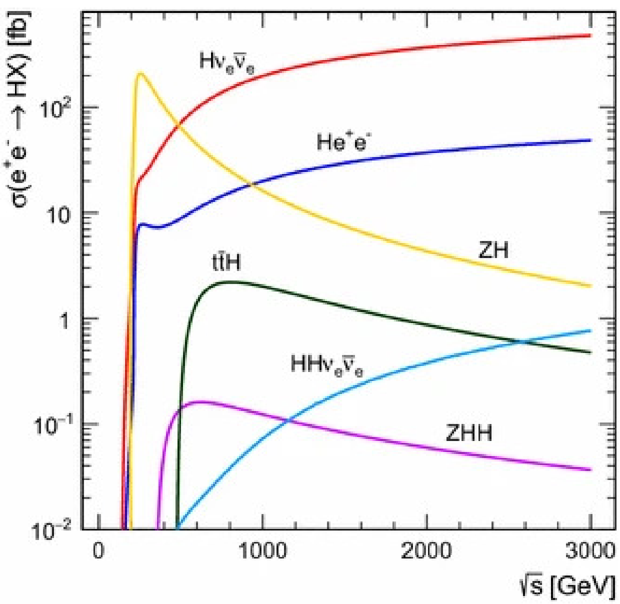

Figure 2: Cross sections of various Higgs boson production channels. At $\sqrt{s}$ = 240 GeV, there is a peak in the ZH process [9]

In this paper, the ZH process is studied carefully, in particular with the hadronization process that occurs in a useful decay mode: Z $\to c\bar{c}$, which has little background for the H $\to b\bar{b}$ decay, which is the most common Higgs decay mode. Jets produced through hadronization are reconstructed, and an analysis of jet properties is conducted to determine the best method of differentiating and identifying candidates for jets produced from Z bosons. Figure 2 illustrates that the ZH process has the highest cross-section (a measure of probability) of Higgs production channels at low energy. In fact, the ZH process has a peak at $\sqrt{s}$ = 240 GeV, which is why FCC-ee plans to operate at that energy during its Higgs factory phase. Figure 3 (below) shows the associated production of a Z boson and Higgs boson occuring through Higgs-strahlung, as well as a Higgs self-coupling interaction.

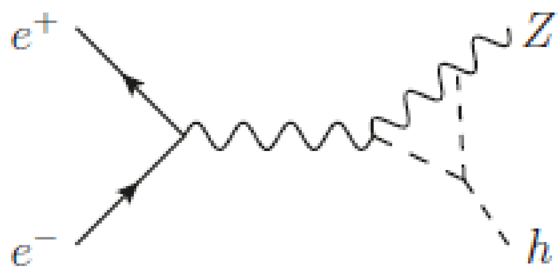

Figure 3: $e^+e^- \to Z$ which then undergoes Higgs-strahlung, resulting in the produced Higgs particles interacting in a self-coupling [10].

This project analyzes a measurement of the ZH cross section in the fully-hadronic final state as one input to a broader fit of many Higgs properties measurements that will constrain the allowed values of the Higgs self-coupling.

2 Hadronization

Of the four fundamental forces, one is of particular interest to this project: the strong force. The strong force is responsible for confining quarks, thus creating hadrons. The confinement of quarks is an especially important property of the strong force, and it comes from asymptotic freedom. The strong force is the only fundamental force that is asymptotically free — i.e. the strength of the force increases as force-carrying particles get farther apart from each other (Figure 4). All of the other fundamental forces become weaker as distance from a source increases.

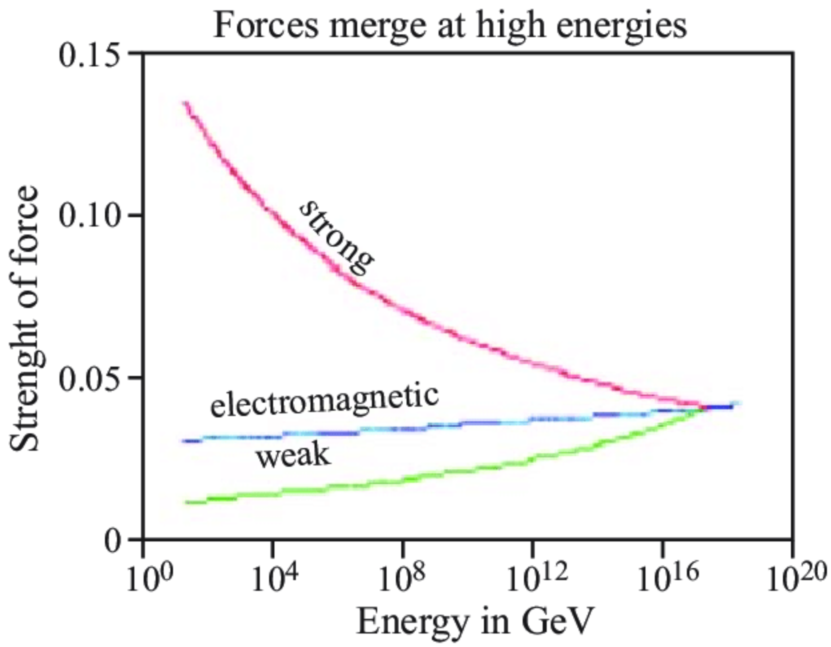

Figure 4: Coupling constants of the fundamental forces vs. energy [11]

Because of asymptotic freedom, quarks are unable to exist alone. However, when high energy particles decay into quarks, enough energy exists in the system for quark pairs to elongate. If the quarks have enough energy, rather than elongating, it will be more energetically favorable for a pair of virtual quarks to be created out of the vacuum, allowing the original quarks to split and pair up with new quarks. This process, called “hadronization,” is constant at high energy particle accelerators. The showers of decay that form due to hadronization are called “jets.”

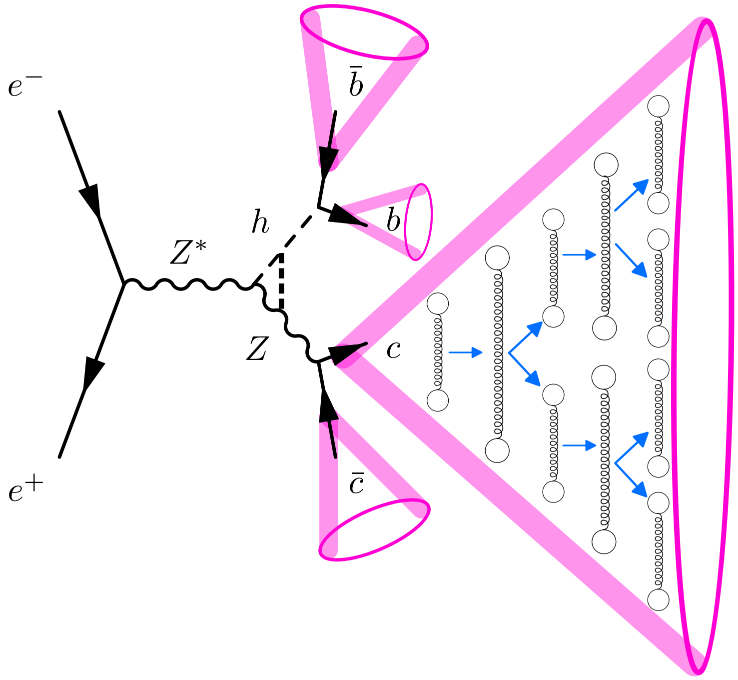

Figure 5: $e^+e^- \rightarrow ZH \rightarrow c\bar{c}b\bar{b}$ Feynman diagram. Jets are depicted by pink cones, and hadronization is emphasized in the c quark jet.

Figure 5 illustrates hadronization in the ZH process. As pictured, after a e^+e^- collision, a Z boson is produced, which then radiates a Higgs bosons through Higgs-strahlung. A virtual Higgs particle is exchanged between the Z and Higgs bosons, thus causing the Higgs self-coupling. The Z boson decays into a charm and anti-charm quark pair, which each hadronize into their own jets. The jet of the charm quark is enlarged for emphasis, bringing attention to the process of virtual quarks being created as quark pairs elongate, creating a shower of hadrons. The Higgs boson decays into a bottom and anti-bottom quark pair, which also hadronize.

The vast majority of particles detected at collider facilities are not primary particles; jets of particles formed through hadronization bombard the detectors. As such, jets must be reconstructed to determine the original process that created them.

3 Methods

Reconstruction of jets is a difficult process because it is difficult to tell if two particles were created from the same primary event. Different reconstruction algorithms have different benefits — the algorithm used in this project is the **Durham k_t algorithm [12]:**

$$d_{ij} = 2\min(E_i^2, E_j^2)(1 - \cos\theta_{ij})$$

for each pair of particles i, j. The i, j pair with the smallest d_{ij} is recombined and all distances are updated. This process continues until either a predetermined number of jets is reached (n_{jets} mode) or all remaining d_{ij} are above some threshold (d_{cut} mode). This project used the Durham algorithm in n_{jets} mode, expecting one jet from each quark produced in the ZH process for a total of four jets: one jet for each of the final state quarks.

Out of the four reconstructed jets, it is assumed that two of the jets are daughters of the Z boson. By identifying Z boson candidates accurately, the recoil mass from these jets can be calculated to identify Higgs boson candidates.

The Durham algorithm was chosen among others because it is exclusive. Exclusive jet reconstruction algorithms prioritize energy recovery by allowing jet radius to be variable, which ensures that every particle belongs to a jet. Energy recovery is prioritized in this project because recoil mass is being considered — if jets are identified as Z candidates but do not contain all of the energy that is actually in the jet, it would lead to a less accurate Z and Higgs mass. Inclusive jet reconstruction algorithms have methods for increasing energy recovery, but none were explored in the scope of this project.

In order to investigate the reconstruction algorithm, data analysis was conducted on Delphes simulations of the IDEA detector in an FCC-like accelerator environment. The simulated observations by the IDEA detector can be seen in Figure 6.

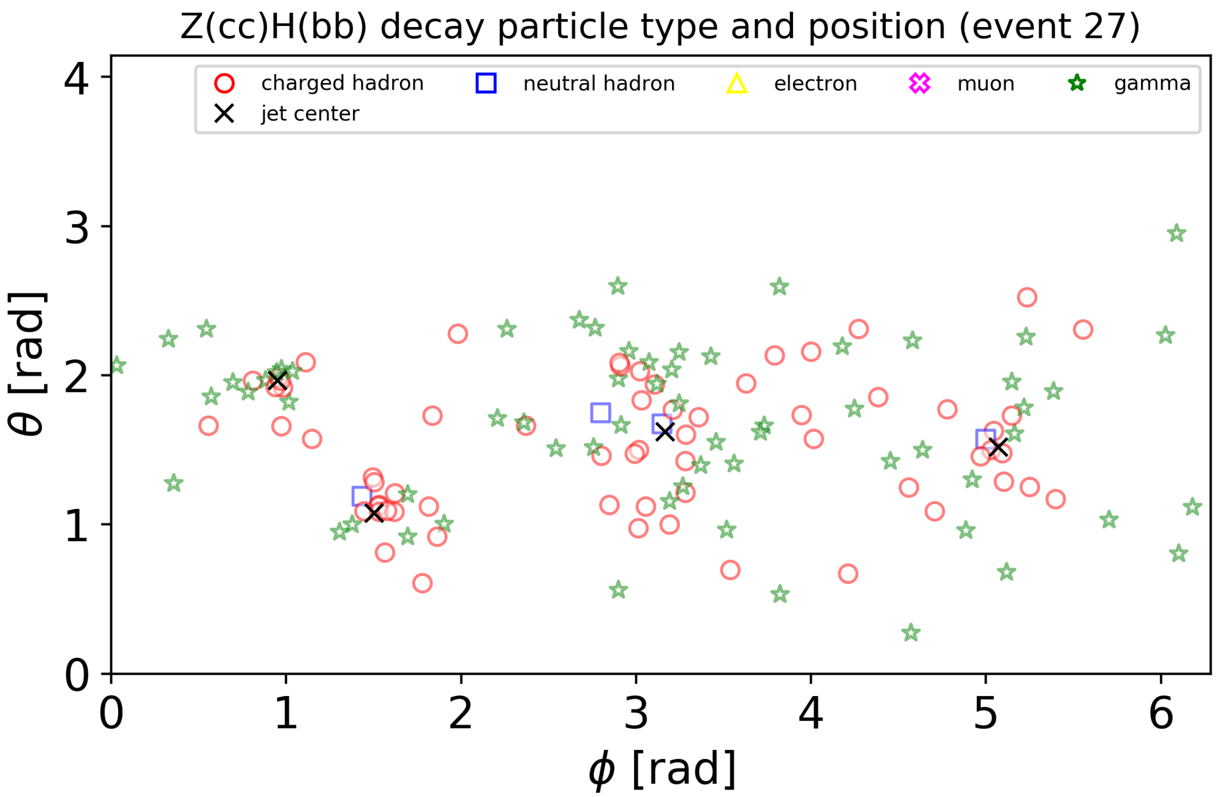

Figure 6: $ZH \to c\bar{c}b\bar{b}$ decay products type graphed over position

Two methods were identified to compare within a ZH process reconstruction: Reconstructing by invariant mass and reconstructing by flavor score.

3.1 Invariant Mass Method

The invariant mass method looks for Z candidates by finding jets that are very close in invariant mass to the Z boson, and then calculates the recoil mass of those jets. For every permutation of two jets from the four possibilities, the mass, energy and momentum of each jet is calculated. Then, the invariant mass for the jet pair is calculated. After iterating over every permutation, the jet pair with invariant mass closest to the mass of the Z boson is chosen: ~91.188 GeV

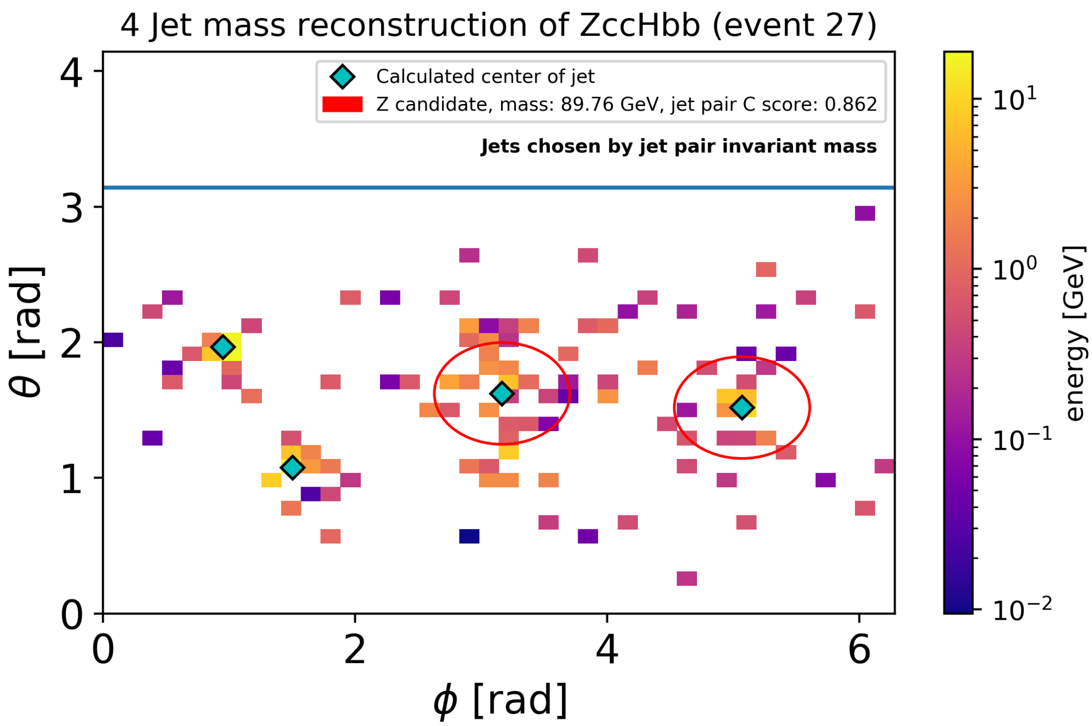

Figure 7: Simulated ZH decay, reconstructed using Durham algorithm. Z candidate jet pair chosen by invariant mass and circled in red (radius not to scale).

3.2 Flavor Score Method

The flavor score method utilizes a neural network approach to flavor tagging jets, where jets are identified by the flavor of the quark that produced it. The neural network is trained to take a jet as input and return a value between 0 and 1 — for example, if a jet has a charm flavor score of 0, it is deemed as 'not charm-like' but a score of 1 is 'charm-like.' Z candidates can be determined by looking for jet pairs with the highest charm flavor values. For every permutation of two jets from the four possibilities, the jet pair with the highest charm flavor score sum is chosen.

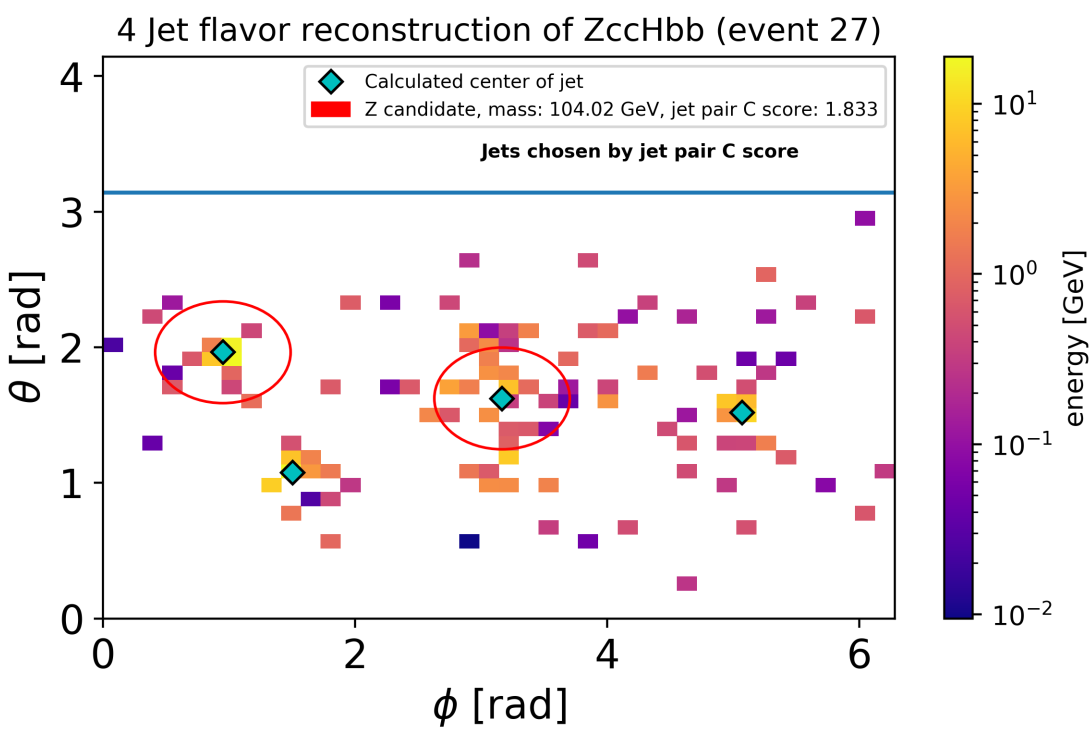

Figure 8: Simulated ZH decay, reconstructed using Durham algorithm. Z candidate jet pair chosen by charm flavor score and circled in red (radius not to scale).

Notice that in the same event, the different methods will choose different jet pairs as Z boson candidates. In order to understand which method may be more efficient, an $\frac{S}{\sqrt{B}}$

analysis is conducted over 100,000 events.

4 Results

4.1 Background and selections

e^+e^- collisions have a number of decay modes which must be accounted for when looking at the ZH process. These modes include $e^+e^- \to ZZ$, $e^+e^- \to Zqq$, and $e^+e^- \to WW$. These processes look similar to the ZH process in detector observations, so they must be separated from the signal.

Background processes occur at much higher rates than the signal process. As such, the simulated results must be scaled accordingly to represent a more physically accurate reality. Table 1 shows the calculations of each scaling factor, calculated by multiplying the cross-section by the luminosity. The luminosity used comes from the current expected luminosity that will be achieved at FCC-ee.

Process	Cross-section (pb)	Luminosity (ab^{-1})	Calculated scaling factor
$ZH \to c\bar{c}b\bar{b}$	0.01359	7.2	97848
ZZ	1.35899	7.2	9784728
Zqq	52.6539	7.2	379108080
WW	16.4385	7.2	118357200

Table 1: Scaling factor values of signal and background processes.

In order to eliminate some of this background, selection cuts were made on both the invariant mass and the flavor scores.

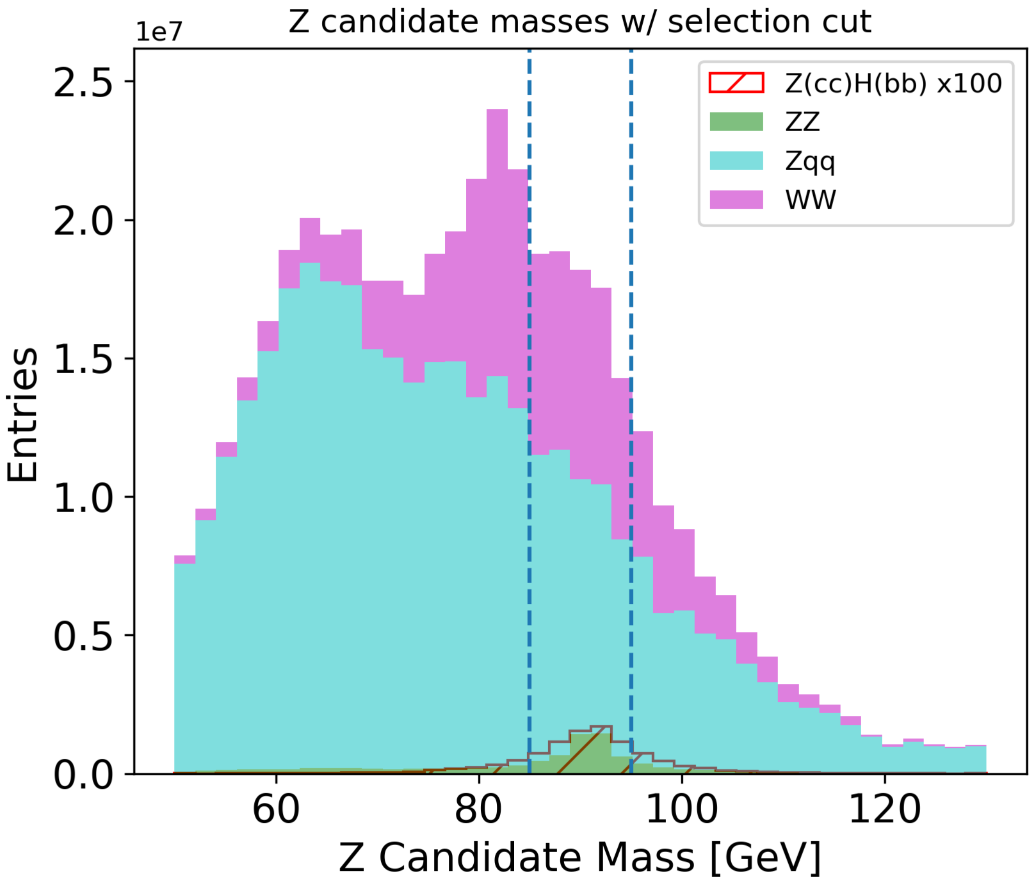

Figure 9: Z candidate masses over 10,000 events. Signal (in red) amplified by a factor of 100 for visibility. Dotted lines at 85 and 95 GeV to indicate selection cut.

Figure 9 illustrates the cut made on jet pairs based on their invariant mass. As seen in the histogram, a significant portion of the background is cut from the data because it lies outside the threshold of 85-95 GeV. This threshold is chosen specifically, by eye, to cut as much background as possible while leaving as much of the signal data (seen in red). Also notice a peak at ~ 80 GeV, which comes from the WW background because the W boson mass is ~ 80 GeV

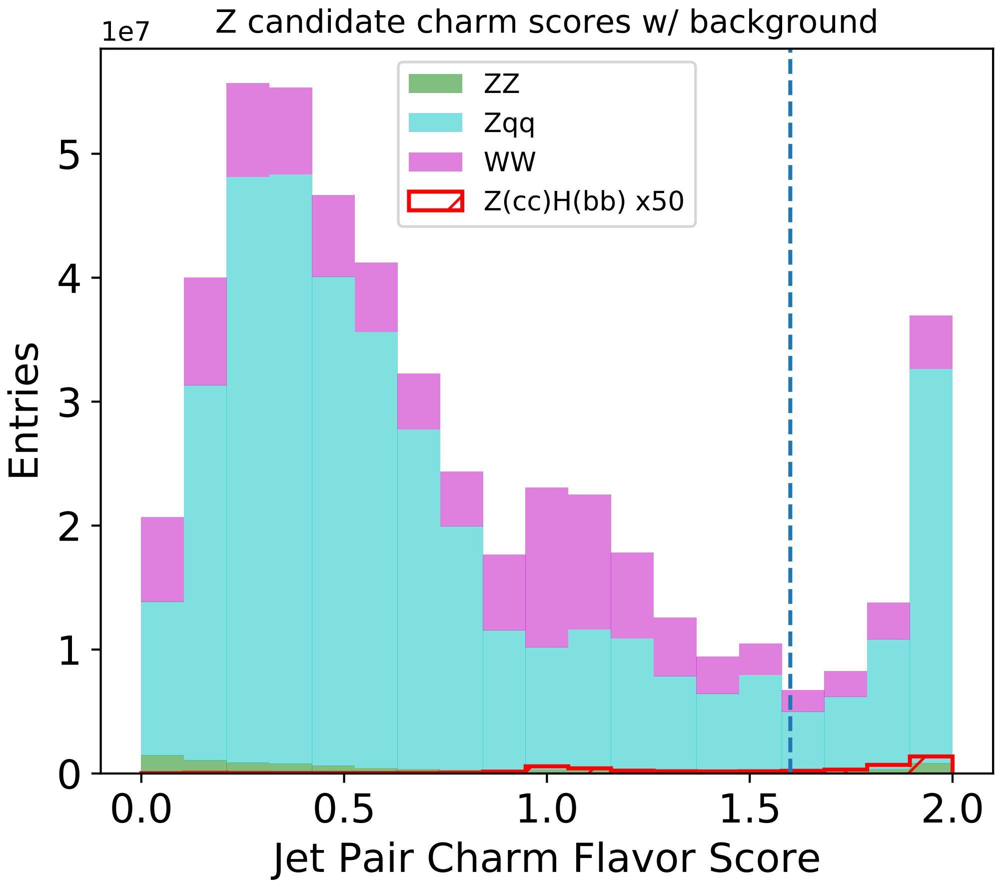

Figure 10: Jet pair charm score (sum) over 100,000 events. Signal (in red) amplified by a factor of 50 for visibility. Dotted line at 1.6 flavor score to indicate selection cut.

Figure 10 depicts the cut made on flavor scores for every jet pair considered. As seen in the histogram, a significant portion of the background is cut from the data because it lies below the threshold of 1.6 total flavor score. This threshold is chosen specifically, by eye, to cut as much background as possible while leaving as much of the signal data (seen in red).

4.2 Invariant Mass Method

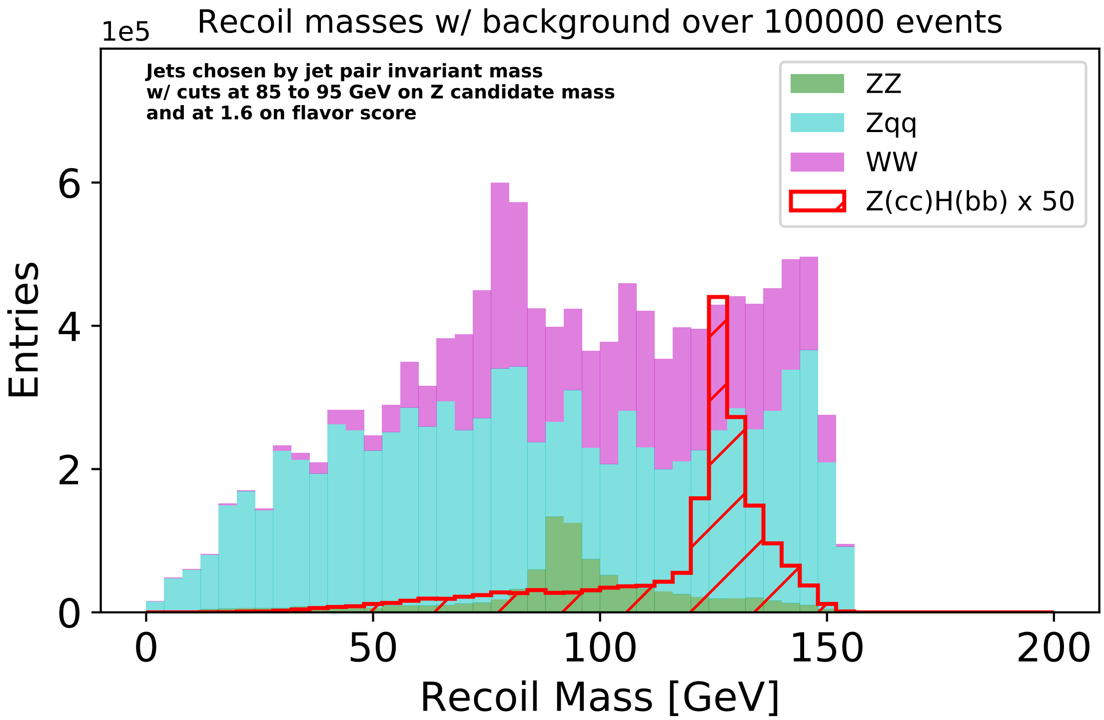

Figure 11: Jet pair recoil masses as chosen by the invariant mass method with selection cuts over 100,000 events. Signal (in red) is amplified by a factor of 50 to emphasize shape.

Figure 11 depicts the calculated recoil masses of jets chosen by the invariant mass method from the signal and background processes, with selection cuts made on invariant mass and flavor score. The signal shows a peak at ∼125 GeV, as expected. Even when amplified by a factor of 50, there is still a large amount of background noise in the 115-135 GeV range, which obscures the signal.

$$\frac{S}{\sqrt{B}} = 14.661$$

4.3 Flavor Score Method

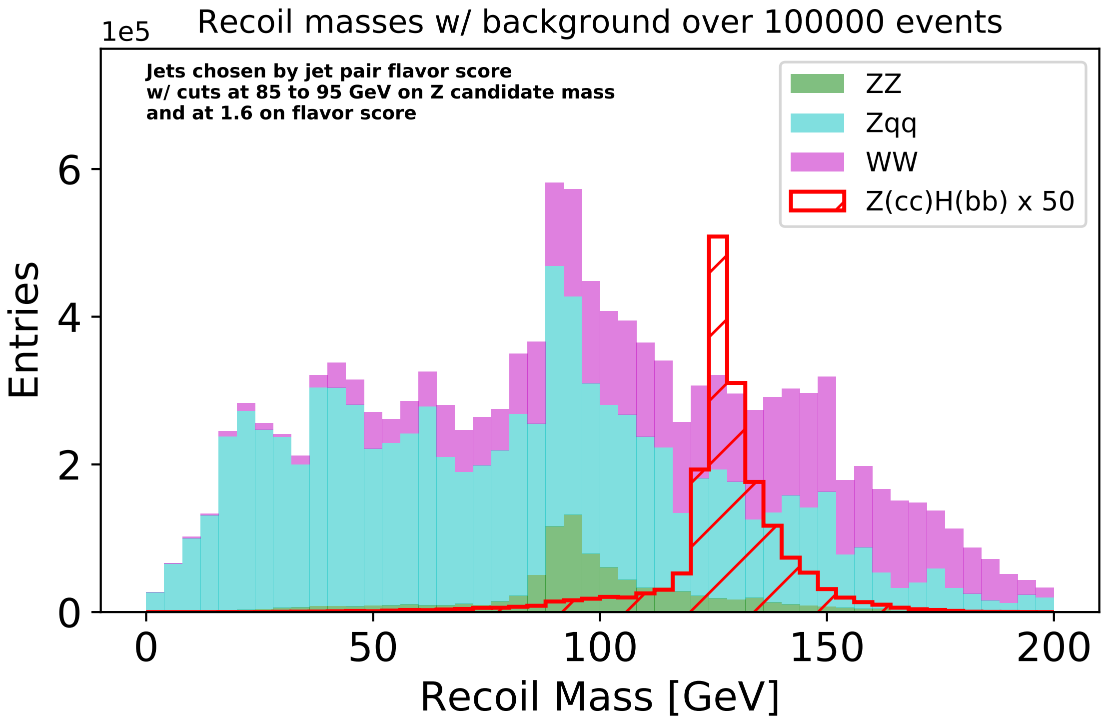

Figure 12: Jet pair recoil masses as chosen by the flavor score method with selection cuts over 100,000 events. Signal (in red) is amplified by a factor of 50 to emphasize shape.

Figure 12 depicts the calculated recoil masses of jets chosen by the flavor score method from the signal and background processes, with selection cuts made on invariant mass and flavor score. The signal shows a peak at ∼125 GeV, as expected. The background shows a noticeable peak at ∼90 GeV, which is explained by the fact that the recoiled particles in the background processes would be particles at roughly that energy (W or Z bosons). There is a sharp fall-off in the background processes near the signal region, allowing the amplified signal to be seen clearly where the amplified signal from the invariant mass is obscured by background.

$$\frac{S}{\sqrt{B}} = 19.959$$

4.4 Discussion

The flavor score method appears to yield a more sensitive analysis of the simulated events as its $\frac{S}{\sqrt{B}}$ is greater than that of the invariant mass method by a factor of $\sim$1.36. This implies that mistakes made when reconstructing jets using the Durham algorithm negatively affect analysis involving the invariant mass of the reconstructed jets more than the flavor tagging neural network. Furthermore, looking for Z candidates by invariant mass will fail on backgrounds that do not have Z bosons at all, such as WW.

5 Conclusions and Future Work

Two methods were identified for tagging jet pairs as Z boson candidates in simulated e^+e^- collisions at the proposed FCC-ee: finding jet pairs with an invariant mass closest to the Z boson mass and finding jet pairs with the highest charm flavor score. The flavor score method has a greater S/sqrt(B) than invariant mass method by a factor of $\sim$1.36. More work must be done to separate the ZH signal from backgrounds to achieve the most sensitive analysis of Higgs self-coupling.

5.1 Future steps

The immediate next step for this project is to get truth information from the original simulation to determine the efficiency of both the invariant mass method and the flavor tagging method. Next, testing other jet reconstruction algorithms, such as inclusive methods or applying different cuts on jet energy, is necessary to better understand how to improve the measurement of Higgs self-coupling. Including other Higgs final states in the signal is also an important step, for example $H \to c\bar{c}$ is currently excluded, as well as many other decay modes. A multivariate analysis would be implemented to maximize $\frac{S}{\sqrt{B}}$ from useful jet properties. Ultimately, this study and others on the ZH process will be used as an input to a fit of many Higgs properties measurements that will constrain the allowed values of the Higgs self-coupling.

6 Acknowledgments

First and foremost, I give special thanks to my mentor, Dr. Abraham Tishelman-Charny for his guidance and patience throughout the course of this project. I would also like to thank to Dr. Elizabeth Brost for her valuable insight and feedback. This project was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). This project is not export controlled.

References

[1] ATLAS collaboration. Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, sep 2012.

[2] CMS collaboration. 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, sep 2012.

[3] J. Ellisa, M. K. Gaillardb, and D. V. Nanopoulosc. An updated historical profile of the higgs boson. *arXiv preprint arXiv:1504.07217*, 2015.

[4] A. Kusenko. Are we on the brink of the higgs abyss? *Physics*, 8:108, 2015.

[5] B. D. Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri. Higgs boson potential at colliders: Status and perspectives. *Reviews in Physics*, 5:100045, nov 2020.

[6] M. C. et al. Higgs physics at the hl-lhc and he-lhc, 2019.

[7] F. collaboration et al. Fcc-ee: the lepton collider: future circular collider conceptual design report volume 2. *European Physical Journal: Special Topics*, 228(2):261–623, 2019.

[8] A. Blondel, P. Janot, N. A. Tehrani, P. Azzi, P. Azzurri, N. Bacchetta, M. Benedikt, F. Blekman, M. Boscolo, M. Dam, et al. Fcc-ee: Your questions answered. *arXiv preprint arXiv:1906.02693*, 2019.

[9] J. De Blas, M. Cepeda, J. D'Hondt, R. Ellis, C. Grojean, B. Heinemann, F. Maltoni, A. Nisati, E. Petit, R. Rattazzi, et al. Higgs boson studies at future particle colliders. *Journal of High Energy Physics*, 2020(1):1–97, 2020.

[10] P. Azzurri, G. Bernardi, S. Braibant, D. d'Enterria, J. Eysermans, P. Janot, A. Li, and E. Perez. A special higgs challenge: measuring the mass and production cross section with ultimate precision at FCC-ee. *The European Physical Journal Plus*, 137(1), dec 2021.

[11] G. Dattoli. The fine structure constant and numerical alchemy, 2010.

[12] M. Cacciari, G. P. Salam, and G. Soyez. Fastjet user manual: (for version 3.0. 2). *The European Physical Journal C*, 72:1–54, 2012.