THE IMPORTANCE OF HIGHLY GRANULAR CALORIMETRY IN RECONSTRUCTING JETS WITH ELECTROMAGNETIC ENERGY AT LINEAR COLLIDERS

by

HARMAN PREET SINGH

A THESIS

Presented to the Department of Physics
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

May 2024

An Abstract of the Thesis of

Harman P. Singh for the degree of Bachelor of Science
in the Department of Physics to be taken June 2024

Title: The Importance of Highly Granular Calorimetry in Reconstructing Jets with Electromagnetic Energy at Linear Colliders

Approved: *James E. Brau, Ph.D.*
Primary Thesis Advisor

The International Linear Collider is the most mature option for a future Higgs Factory capable of probing new physics beyond the Standard Model. Precision measurements will be crucial to its mission, which will be achieved through exceptionally accurate detectors like SiD. An upgrade has been proposed for the design of the Electromagnetic calorimeter, replacing the current analog pixel model with a higher granularity one with digital pixels. This analysis considers the potential improvements to physics studies with this upgrade by considering the capabilities of separating dijet events from W and Z decays with the current detector model and potential improvements with a high granularity upgrade.

Specifically, this study identifies 10 physics based separation parameters to train and utilize a neural network to perform the separation. Importantly, the separation of W and Z events will be crucial to the goals of a future linear collider since W pair production is an overwhelming source of background to Higgs-strahlung events. The Higgs-strahlung channel will be crucial for beyond Standard Model searches since they provide a clean and consistent channel to probe the invisible decay of the Higgs which could include new particles like dark matter.

Acknowledgements

I would like to thank the members of the University of Oregon SiD group who have made this thesis possible including Professor James Brau, Dr. Chris Potter, and Bayley Burke. I want to give special thanks to Jim, who has been an amazing advisor by providing me with unique opportunities to explore my interests in physics, guiding me through all my confusion, and pushing me to learn more and improve my skills as a physicist. I appreciate Chris's contribution to this thesis by providing the data for my analysis and lending critical advice for catching countless bugs.

I'm thankful to every University of Oregon physics faculty member who has taught me and encouraged my passion and curiosity in physics which has led to me pursuing a PhD in theoretical particle physics. I would like to extend special thanks to Professor Spencer Chang, Professor Graham Kribs, Professor Tien-Tien Yu, and Professor Laura Jeanty whose mentorship and instruction has fostered my pursuit of physics.

Finally, I want to thank and dedicate this thesis to my parents, Davinder Kaur and Gagandeep Singh. Their continuous sacrifice and love have not only made this degree and thesis possible but provided me with the invaluable opportunity to pursue my dreams. I'm grateful to be their son since they have taught me the importance of family, culture, Sikhi, and education. I would not be who I am without them and cannot wait to celebrate with them. Last but not least, I would like to thank the best dog known to humanity, Mac Singh, for his undeniable cuteness and the unlimited joy he has brought into my life.

Table of Contents

The Standard Model and Beyond	8
The International Linear Collider and SiD	11
W and Z Dijet Events	16
Separation Parameters	23
Training and Evaluating a Separator	40
Post-Separation Analysis	53
Conclusion	60
References	62

List of Figures

Figure 1: The Standard Model and its Interactions 8
Figure 2: The International Linear Collider 11
Figure 3: The SiD 13
Figure 4: TDR vs MAPS 14
Figure 5: Typical Particle Interactions in a Detector 15
Figure 6: Primary Feynman Diagrams of Higgs Production at the ILC 17
Figure 7: Cross Section of the Primary Higgs Production Processes vs Center of Mass Energy 17
Figure 8: Comparison of Dijet Invariant Mass of W and Z Events 18
Figure 9: Dijet Event Display 21
Figure 10: Distribution of Missing Particle Tracks from PFO Data 24
Figure 11: Comparison of Dijet Recoil Mass of W and Z Events 25
Figure 12: Distribution of Summed Neutrino Event Energy 25
Figure 13a: Comparison of Energy of Jet 1 of W and Z Events 26
Figure 13b: Comparison of Energy of Jet 2 of W and Z Events 27
Figure 14a: Cross Section of SM Processes at the ILC with eLpR Polarization 28
Figure 14b: Cross Section of SM Processes at the ILC with eRpL Polarization 28
Figure 15: Comparison of Dijet Beam Angle of W and Z Events 29
Figure 16a: Comparison of Jet 1 b Count of W and Z Events 32
Figure 16b: Comparison of Jet 2 b Count of W and Z Events 32
Figure 16c: Comparison of Jet 1 c Count of W and Z Events 32
Figure 16d: Comparison of Jet 2 c Count of W and Z Events 33
Figure 17: Comparison of Dijet Opening Angle of W and Z Events 34
Figure 18a: Post-Cut Comparison of Dijet Invariant Mass of W and Z Events 35
Figure 18b: Post-Cut Comparison of Dijet Recoil Mass of W and Z Events 35
Figure 18c: Post-Cut Comparison of Jet 1 Energy of W and Z Events 36
Figure 18d: Post-Cut Comparison of Jet 2 Energy of W and Z Events 36
Figure 18e: Post-Cut Comparison of Dijet Beam Angle of W and Z Events 37
Figure 18f: Post-Cut Comparison of Dijet Opening Angle of W and Z Events 37
Figure 18g: Post-Cut Comparison of Jet 1 b Count of W and Z Events 38
Figure 18h: Post-Cut Comparison of Jet 2 b Count of W and Z Events 38
Figure 18i: Post-Cut Comparison of Jet 1 c Count of W and Z Events 38

Figure 18j: Post-Cut Comparison of Jet 2 c Count of W and Z Events 39
Figure 19: ROC Curve for Training a W and Z Classifier 41
Figure 20a: Impact of W and Z Dijet Invariant Mass on Training 42
Figure 20b: Impact of W and Z Dijet Recoil Mass on Training 42
Figure 20c: Impact of Jet 1 Energy of W and Z Events on Training 43
Figure 20d: Impact of Jet 2 Energy of W and Z Events on Training 43
Figure 20e: Impact of W and Z Dijet Beam Angle on Training 44
Figure 20f: Impact of W and Z Dijet Opening Angle on Training 44
Figure 20g: Impact of Jet 1 b Count of W and Z Events on Training 45
Figure 20h: Impact of Jet 2 b Count of W and Z Events on Training 45
Figure 20i: Impact of Jet 1 c Count of W and Z Events on Training 45
Figure 20j: Impact of Jet 2 c Count of W and Z Events on Training 46
Figure 21a: Neural Network Weight and Signal Efficiency 47
Figure 21b: Neural Network Weight and Signal Efficiency 47
Figure 22a: Post-Separation Signal Mass Distribution 48
Figure 22b: Post-Separation Signal Recoil Mass Distribution 48
Figure 22c: Post-Separation Signal Jet 1 Energy Distribution 49
Figure 22d: Post-Separation Signal Jet 2 Energy Distribution 49
Figure 22e: Post-Separation Signal Dijet Beam Angle Distribution 50
Figure 22f: Post-Separation Signal Dijet Opening Angle Distribution 50
Figure 22g: Post-Separation Signal Jet 1 b Count Distribution 51
Figure 22h: Post-Separation Signal Jet 2 b Count Distribution 51
Figure 22i: Post-Separation Signal Jet 1 c Count Distribution 51
Figure 22j: Post- Separation Signal Jet 2 c Count Distribution 52
Figure 23a: Lepton Angle and Energy Phase Space for Signal ≥ 0.97 54
Figure 23b: Lepton Angle and Energy Phase Space for Signal ≥ 0.90 55
Figure 24a: Jet 1 Energy Distribution of Z Events 56
Figure 24b: Jet 1 Energy Distribution of W Events 56
Figure 25: Scaled Jet Energy Resolution 56
Figure 26a: Jet Energy in Photons 58
Figure 26b: Jet Energy in Leptons 58
Figure 26c: Jet Energy in Neutral Hadrons 58
Figure 26d: Jet Energy in Charged Hadrons 59

Figure 27: Scaled Jet Energy Resolution and Jet Energy in Photons 59
Figure 28: Invariant Dijet Mass Distribution Post Leptonic Energy Cut 60

The Standard Model and Beyond

The Standard Model (SM) is a beautiful theory at the heart of modern particle physics that explains and classifies all elementary particles and the interactions of three out of the four fundamental forces as seen in Figure 1. At the center of the SM, is a set of symmetries that explain the fundamental dynamics of nature which determine the interactions of particle constituents in the SM. The SM is written in the mathematical language of quantum field theory wherein one considers particles as excited momentum states of fundamental fields that span all spacetime. In this framework, a particle's internal angular momentum or spin provides a direct method of classifying particles. Whole integer spin particles or bosons like the photon, gluon, $W^{\pm}$, and Z^0 bosons are force carriers. Half-integer spin particles or fermions are the building blocks of matter.

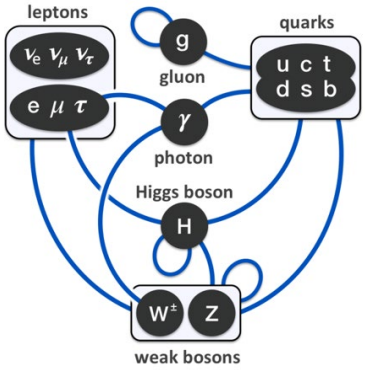

Figure 1: The Standard Model and its Interactions

This diagram shows the content of the SM and their interactions. The interactions can be described by two symmetry gauge groups. Specifically, electroweak interactions are symmetric under $SU(2)_L \times U(1)_Y$ so couplings are tied to chirality and hypercharge. Strong interaction are symmetric under $SU(3)_C$ so the interactions are based on color charge https://en.wikipedia.org/wiki/File:Elementary_particle_interactions_in_the_Standard_Model.png

Fermions can be classified further into leptons and quarks based on their potential interactions. Leptons importantly don't interact with the strong force and come in three

generations which are referred to as electronic, muonic, and tauonic leptons. Each generation contains an electromagnetically charged particle and an extremely light electromagnetic neutral particle called a neutrino. On the other hand, quarks couple to the strong force and are charged particles that come in different flavors (up, down, charm, strange, top, and bottom). Particles that interact under the strong force will gain a new type of charge called color charge, leading to a unique phenomenon known as color confinement wherein a quark can never be isolated. Consequently, quarks form composite particles like baryons and mesons. Baryons like neutrons and protons are constructed from a quark triplet while mesons like pions, and kaons are constructed from a quark doublet.

Most bosons serve critical dynamical roles since they become force carrying particles guiding potential interactions. Photons are massless particles that mediate the electromagnetic force responsible for light and interactions among electrically charged particles. Gluons are also massless particles, but they mediate the strong force responsible for gluing quarks together to form hadrons. Unlike the photon, gluons carry the charge for the strong interaction, color, which means they interact with one another leading to strong interactions growing with distance like an elastic string. The charged W bosons (W^+/W^-) and the neutral Z boson are massive bosons that mediate the weak force which is responsible for radioactive phenomena like beta decay.

One can consider these forces to exist independently, but this provides an incomplete picture. As James Clerk Maxwell showed electricity and magnetism were the same force at higher energies, electromagnetism and the weak force unify into the electroweak force at higher energies. Given the photon, W bosons, and Z boson are carriers of the singular electroweak force, a mechanism is required to explain the varying masses of the boson. The Higgs boson was the final piece of the puzzle to the SM explaining how the symmetries of the electroweak force

are spontaneously broken into electromagnetism and the weak force. In the process, it explained how fundamental particles gained mass through their interaction with the Higgs field.

The 2012 discovery of the Higgs boson at the Large Hadron Collider (LHC) highlighted the vast experimental success of the SM [1]. However, its discovery brought a plethora of questions. Some questions were concerned with its measured properties like its invariant mass measured around 125 GeV/c^2 which is on the lower end of the predicted mass range. Other questions surrounding its predicted coupling strengths to specific particles like itself remain unanswered due to experimental capabilities. Additional questions arise from probing potential properties of the Higgs such as potential composite behavior and multiple Higgs. This final set of questions lies in the realm of probing physics beyond the SM (BSM). Regardless of its experimental and theoretical success, the SM is an incomplete theory since it lacks a description of key fundamental realities of our universe such as gravity, grand unification, matter-antimatter asymmetry, and dark matter which composes ~80% of the matter in the universe. To probe these questions about the Higgs and BSM physics, an exceptionally accurate experiment capable of probing higher energies and being able to function as a Higgs factory is needed.

The International Linear Collider and SiD

The International Linear Collider (ILC) is a proposed 31 km linear electron-positron collider as seen in Figure 2 and is the most mature option for a future Higgs factory [2]. Importantly, the ILC will have polarized electron-positron beams which provide a significant advantage for precision physics over proton-proton colliders like the LHC. Crucially, electrons are fundamental particles meaning they collide with the entirety of the center of mass energy and produce clean events. Moreover, these beams are polarized by stripping electrons from a cathode and undulating this original electron beam to produce the positron beam. In this context, polarization refers to the alignment of a particle's spin with its direction of motion. Left-handed particles are anti-aligned while right-handed particles are aligned.

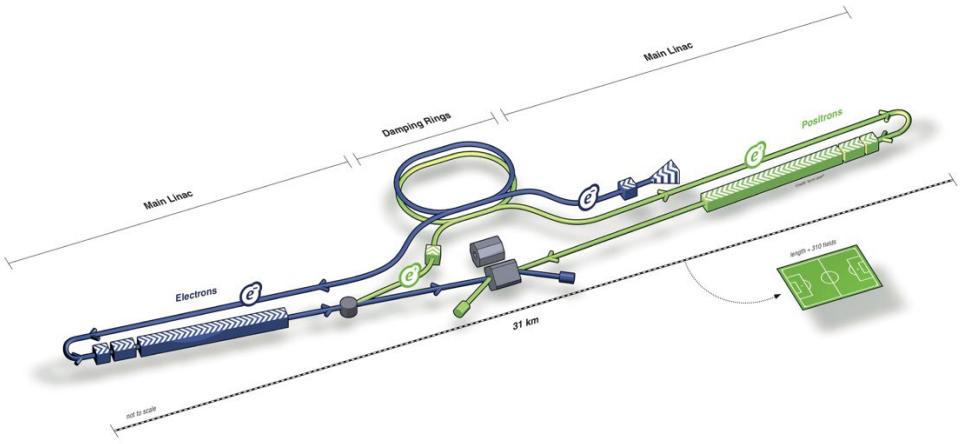

Figure 2: The International Linear Collider

This figure shows the current general design of the ILC. The electron beam is shown in blue, and the positron beam is shown in green, and it is critically being shown being produced from the electron beam. For a sense of scale, every dot in the dotted line is roughly the size of a soccer field [2].

The current ILC design for a center of mass energy of 250 GeV is projected to achieve an alternating polarization of 80% electrons of a particular handedness and 30% positrons of a particular handedness over the 4 potential polarization states. The plan includes dividing the

frequency of the 4 states into 5% as $e_L^- e_L^+$ (eLpL), 45% as $e_L^- e_R^+$ (eLpR), 45% as $e_R^- e_L^+$ (eRpL), and 5% as $e_R^- e_R^+$ (eRpR). [3] These polarizations will be crucial for probing electroweak interactions since they are parity-violating processes. For example, the W bosons maximally violate parity since they exclusively couple to left-handed particles and right-handed anti-particles. Finally, the linear geometry of the ILC avoids energy loss from synchrotron radiation which limit circular colliders. Synchrotron radiation becomes a larger issue for colliders with electron beams since the energy of a charged accelerating particle loses, increases with smaller mass (mass dependence of $\sim m^{-4}$). This enables the ILC to probe an initial center of mass energy of 250 GeV and future energies of 500 GeV, 1 TeV, and beyond. Importantly, linear colliders like the ILC will dump their beams while circular colliders will recycle their beams to get multiple collisions. Linear colliders maintain a competitive luminosity by focusing their beams to smaller points.

Due to the linear geometry, the ILC's detectors, referred to as SiD and ILD (International Large Detector), share an interaction point that requires a push and pull method to perform measurements. Critical to the physics goals of the ILC are precision measurements which motivate the design efforts. SiD achieves these requirements by combining results from the silicon tracking system, silicon-tungsten Electromagnetic Calorimeter (ECal), and Hadronic Calorimeter (HCal) in conjunction with a Particle Flow Algorithm (PFA) to provide highly precise jet and particle energy measurements [4]. The current design as seen in Figure 3, is the proposed design in the Technical Design Report (TDR) where the ECal is composed of alternating layers of high-Z tungsten and narrow gaps for silicon, which results in a small Molière radius and narrow showers. Energy measurements are performed by 13 mm^2 analog pixels.

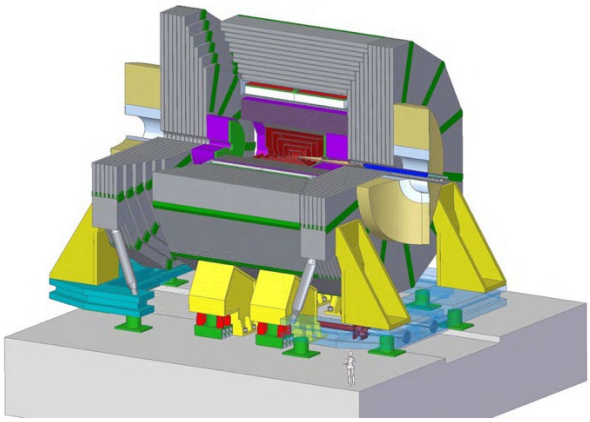

Figure 3: The SiD

This is a 3-dimensional depiction of the SiD where the inner parts of the detector show the HCal in magenta, ECal in green, and the tracker in red. For scale, a person is depicted next to the detector [4].

However, a higher granularity model of the ECal has been proposed where the pixels are replaced with digital Monolithic Active Pixel Sensors (MAPS) of size 25 μm × 100 μm. The MAPS design utilizes digital pixels that record coordinates for energy measured over a threshold, without a measure of its amplitude. Moreover, studies have shown this higher granularity ECal can recover the precision of the TDR design in terms of energy measurements and has the potential to improve through the usage of a weighted clustering algorithm [5]. Crucially, MAPS provides increased spatial separation of detected particles as seen in Figure 4.

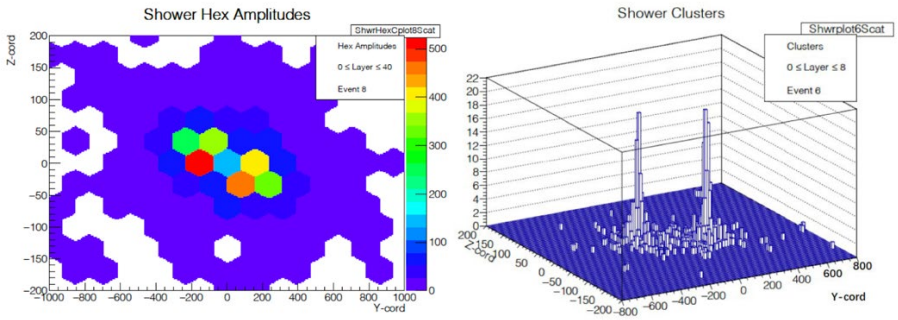

Figure 4: TDR vs MAPS

This figure depicts event displays of two 10 GeV electron showers separated by 1 cm using the TDR design (left) and the MAPS design (right). The increased granularity of the MAPS design clearly shows an increased distinguishability of the showers, while the TDR design shows some confusion about which shower is which and will only worsen as the shower separation is shortened [5].

MAPS should improve the ability to separate different particles from neighboring energy deposits in tightly dense regions due to the fine granularity, particularly for electromagnetic showers, which are narrow due to the high-Z tungsten structure. This improved precision should greatly aid the PFA, which uses a holistic approach to combine measurements from the tracker, ECal, and HCal to differentiate detected objects as charged particles (electrons, and charged hadrons), neutral hadrons (neutrons, neutral kaons, etc.), and other neutral particles (photons) [6]. Charged particles will curve a measurable amount in the presence of the magnetic field which can be detected using the tracker. Charged particles like leptons and charged hadrons are differentiated since the leptons like electrons will deposit their energy in the ECal and the hadrons mostly in the HCal. On the other hand, neutral particles like photons and neutral hadrons are differentiated the same way but since they do not curve in magnetic fields, they will be invisible to the tracker. These different classes of particles will each have their own interactions within the detector as seen in Figure 5. However, quantifying the MAPS model's improvement when reconstructing particles and jets needs further analysis. A critical test to evaluate the

detector's ability to perform precision measurements is the separation of dijet events from W and Z decay.

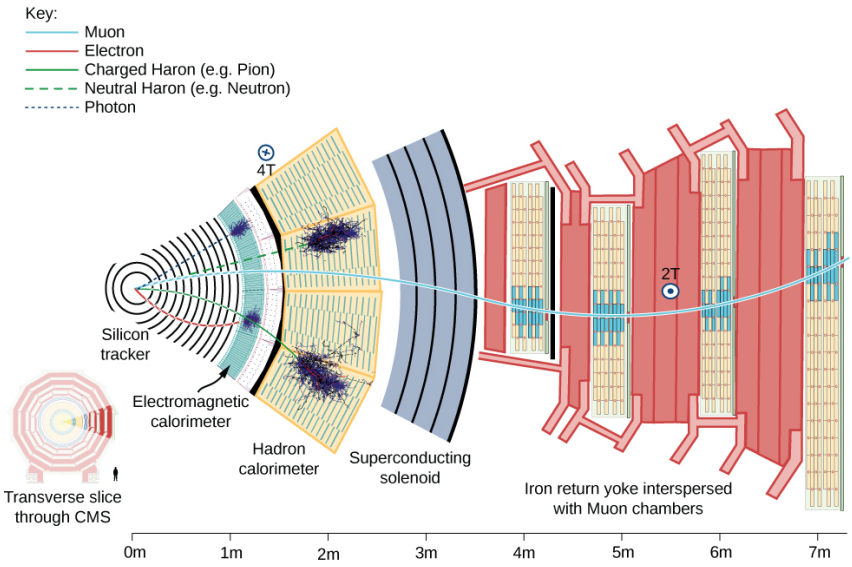

Figure 5: Typical Particle Interactions in a Detector

This figure depicts how various particles interact within the Compact Muon Solenoid (CMS) Detector at the LHC. Even though SiD is distinct from CMS, we know particles will interact similarly. Crucially, the figure visually shows how the different particle types can be differentiated [https://pressbooks.online.ucf.edu/osuniversityphysics3/chapter/particle-accelerators-and-detectors/].

W and Z Dijet Events

Jet studies are crucial in studying the accuracy of detectors since jets are clustered collimated showers of particles resulting from the hadronization of quarks and gluons. The precise reconstruction of jets is a key goal of SiD since the internal structure can illustrate the interactions of particles within the jets and what happened directly following the initial collision [7]. Additionally, precision in jet measurements will be integral to measuring Higgs events. The three primary Higgs production processes (Higgs-strahlung, W fusion, and Z fusion) are shown in Figure 6. Furthermore, Figure 7 shows that Higgs-strahlung is the dominant process at 250 GeV. Importantly, the Higgs-strahlung channel will be critical to BSM studies since it will be able to probe exotic decay channels of the Higgs including invisible decays. In the SM, the Higgs is predicted to decay invisibly (neutrinos) around 0.1% of the time, but specific extensions to the SM can vary this rate [8]. Even if the Higgs decays invisibly, energy-momentum preservation allows for measurements of the Higgs using measurements from the same event.

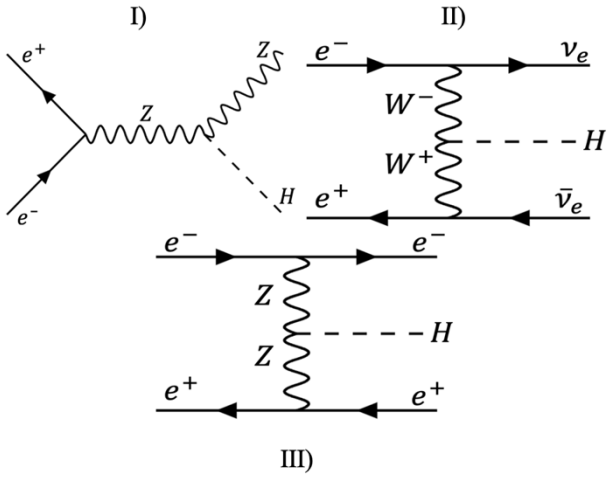

Figure 6: Primary Feynman Diagrams of Higgs Production at the ILC

This figure shows the Feynman diagrams for the main processes of Higgs Production at an e^+e^- collider. Importantly, these diagrams visually show the perturbative approach of finding the amplitude of transitioning from some initial state to a final state in quantum field theory in momentum space. The first diagram shows the Higgs-strahlung. The second and third diagrams are the vector boson fusion processes since they involved the respective vector gauge bosons that mediate the weak force.

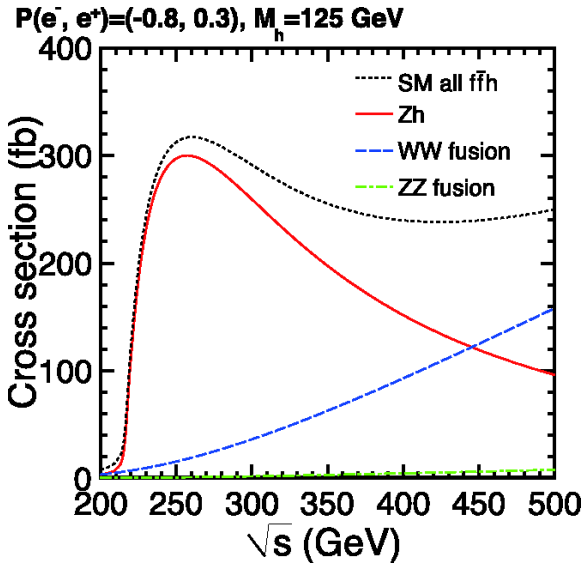

Figure 7: Cross Section of the Primary Higgs Production Processes vs Center of Mass Energy

This figure shows the cross section of Higgs-strahlung, W fusion, and Z fusion as a function of the center of mass energy. Within this context, the cross section gives the probability a certain process will occur due to a specific particle interaction or collision in units of barns where $1\text{b} = 10^{-28}\text{m}^2$ [9].

In the case of Higgs-strahlung and invisible Higgs decay, the recoil off the reconstructed Z can be used to indirectly measure the Higgs. However, the proximity of the Z boson mass (~91 GeV/c^2) and W boson mass (~80 GeV/c^2) demands exceptionally accurate jet energy measurements to reconstruct the originating particle's invariant mass to effectively discriminate between Z jets and W jets [10]. The proximity of the Z and W mass is further emphasized through the inclusion of their decay width, which means a boson can be produced within a 2-2.5 GeV/c^2 range of their invariant mass. Also, missing energy from jets or a third jet can widen the measured mass curve, which can make it more difficult to tag Z and W jets as seen in Figure 8. Improving jet reconstruction and energy measurements through increasing the granularity of the ECal has been proven in previous studies using simulations of multi-TeV jets at hadronic colliders, which enhanced the effectiveness of tagging Z and W jets. However, a dedicated study for linear colliders like the ILC and its detectors like the SiD using lower energy jets (< 100 GeV) has the potential to show the same results [11].

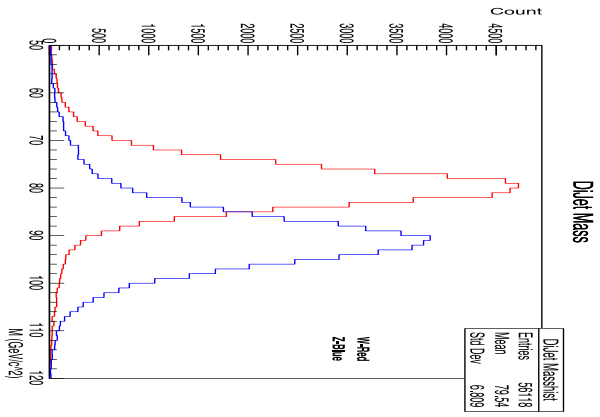

Figure 8: Comparison of Dijet Invariant Mass of W and Z Events

This plot shows the invariant mass of the combined dijet system using PFO data. W is plotted in red while Z is plotted in blue. The plot displays an equal number of W and Z events.

Since the ILC has not been built, the entirety of the data that has been analyzed is sourced from simulations. Particularly, Dr. Chris Potter produced 100,000 Higgs-strahlung and 1,000,000

W pair production events using Monte Carlo (MC) simulations that accurately simulate electron-positron collisions in the environment of the SiD, which produce truth values to compare against and provide a quantifiable sense of the detector's accuracy [12]. MC simulation software like WHIZARD and MADGRAPH were used to accurately replicate realistic physics, model the probabilistic nature of particle physics, and take into account linear leptonic collider effects like beamstrauhlung [13] [14]. To focus on jet events, PYTHIA was used in conjunction with MC simulation software to replicate realistic hadronization using string models [15]. Additionally, GEANT4 was used to simulate response of generated particles with the detector [16].

The output of these simulations is then reconstructed by a PFA like Pandora, which replicates these events in a manner that reflects how the detector would measure these events and produce realistic results [17]. The output of the PFA will be in the form of Particle Flow Objects (PFOs), which represent measured particles, and the MC simulations will output MC particles, which represent the truth. Output particle types will be tagged with an identifying integer using the Particle Data Group's (PDG) particle numbering scheme (anti-particle associated with negative integers) [10]. Alongside the ID number, the particle's energy and momentum will be listed as an energy-momentum four vector. The simulations were run using a center of mass energy of 250 GeV, which is the planned initial running energy of the ILC.

To analyze the reconstruction of the Z boson using jets, we exclusively considered Higgs-strahlung events where the Higgs was forced to decay invisibly and the Z decaying into a quark anti-quark pair which occurs at a frequency of ~65% for the Z [10]. The invisible decay (neutrinos) of the Higgs simplifies the analysis to only two jet events and probes the measurement sensitivity of this decay. Neutrinos rarely interact with matter, so it is a safe assumption to consider them invisible for our analysis. To focus on jet production, we utilized

events where the Z boson decays into a quark and anti-quark pair, which can be easily identified since they will contain substantially more particles than other events, but we can verify the Z decayed into particles with an associated ID number of ±1, ±2, ±3, ±4, and ±5, which represent the different flavored quarks/anti-quarks (except the top quark since it is too heavy). Approximately 65,000 of the simulated Z events were $Z \rightarrow q\bar{q}$ events. Critically, the Z decay processes will include final state neutrinos which will be measured as missing energy for the Z events.

Similarly, we exclusively considered pair produced W events where the original electron-positron pair produced a W^+ and W^- bosons. To simplify the analysis to two jet events and make these events comparable to the Z events, the W^+ was chosen to decay into a quark and anti-quark pair which occurs at a frequency of around 68% [10]. The W^- on the other hand, decays into an anti-muon neutrino and muon which can radiate photons that are removed for the dijet formation portion of the analysis. Importantly, all 1 million W events were forced to decay in this manner.

After producing and collecting events for jet analysis, we will form jets using PyJet, which is a Python-based implementation of FastJet allowing for the usage of NumPy [18] [19]. Importantly, the jet reconstruction algorithm can become somewhat simplified by acknowledging the ILC is a linear collider meaning, the algorithm can rely on energy and spherical coordinates instead of transverse momentum and pseudorapidity [20]. We use the preset generalized k_t algorithm for e^+e^- collisions where we specify the general parameters p (style of clustering) and R (opening half angle). We choose $p = -1$ to simplify the algorithm to the anti-k_t algorithm so it produces more cone-like clusters, while setting $R = 1$ emphasizes the

fact we are considering relatively widespread jets. This analysis will focus on events where the energy of the event is primarily within the highest energy jets as seen in Figure 9.

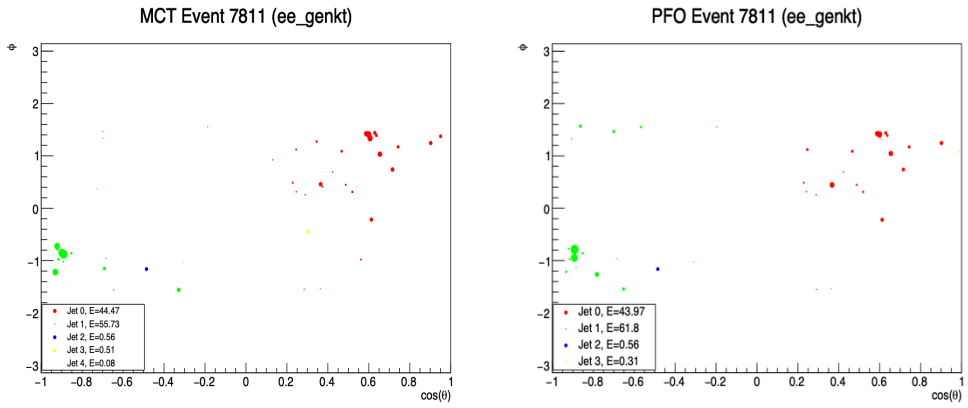

Figure 9: Dijet Event Display

This figure shows an example of a dijet event that we will consider. The left represents the simulated truth (MCT) while the right represents the reconstruction (PFO). Each dot represents a particle where the size of the dot corresponds to its energy and the color represents its clustered jet.

To identify well-measured dijet events for this analysis we will implement statistical cuts to improve the analysis. To ensure we evaluate only dijet events, a preliminary cut was made to exclusively consider events where the two highest energy jets composed ⅔ of the entire event's energy. To ensure we focus on "well-measured" events we match MC jets to their respective PFO jets using the metric

$$Y = \sum_{\mu=0}^{3} \left(1 - \frac{p^{\mu}_{\text{PFO}}}{p^{\mu}_{\text{MCT}}}\right)^2, \ (\mu \text{ runs over spacetime indices}). \tag{1}$$

Since any given match isn't necessarily a good match indicative of accurate reconstruction, a cut of $\log(Y) < 0$ is imposed to identify well-measured jets. If two MCT jets and two PFO jets survive this cut, they will be considered for analysis. Since we expect these events to be dijet events, the remaining particles within the event are summed into the nearest dijet by finding the

smallest angle formed by the particle's momentum vector and one of the two jet axes. Following these cuts, we are left with ~85% of the Z events and ~76% of the W events.

Separation Parameters

An effective method to separate W and Z events will need to rely upon the physical properties of the bosons which can be constructed from the dijet system. Specifically, various separation parameters need to be identified to train a neural network to evaluate whether a particular event is from a signal event (Z) or a background event (W). One can immediately see the bosons have different electromagnetic charges and masses which should provide critical separation parameters; however, the mass was shown to be a crucial region of confusion since the shoulders of their mass distributions significantly overlap as seen in Figure 8. Even though mass cannot exclusively be used as a separation parameter, it can still provide critical separation information and provide a metric for a pre-training cut on the data to remove background W events. On the other hand, the charge of these events corresponds to the combined charge of the final state particles which can differ quite a bit due to missing tracks as seen in Figure 10. These missing tracks make electromagnetic charge an unreliable parameter since the W has a charge of 1 and the Z has a charge of 0 so slight deviations from these values can confuse a separator.

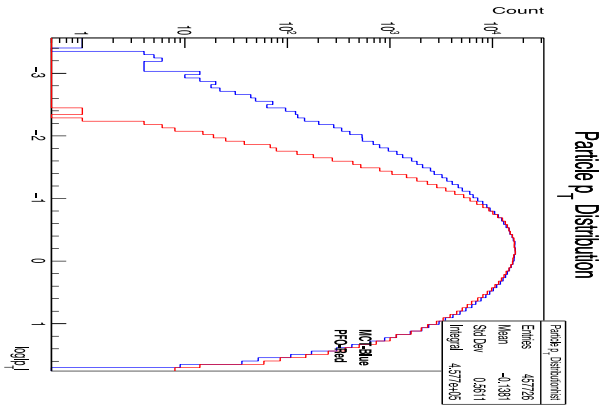

Figure 10: Distribution of Missing Particle Tracks from PFO Data

This semi-log distribution plots the logarithm of the transverse momentum defined as $p_T = |\vec{p}|\sin(\theta_B)$ where θ_B is the angle formed by the momentum vector and the beam axis. The reconstructed data is plotted in red and the simulated data in blue. The distribution shows the PFO data missing low p_T particles since they do not curve as much as needed in the tracker.

In addition to their fundamental properties, the unique couplings of the W and Z to fermions and kinematics of the Higgs-strahlung and W pair production process can be the basis of additional separation parameters. The unique kinematics of each process provide 3 critical separation parameters. Importantly, the recoil mass provides a cleaner separation between W and Z events than their invariant mass as seen in Figure 11. The recoil mass is calculated using the principle of energy-momentum conservation. Since the original energy of the colliding electron-positron system is known, one can use the measured energy to find the missing energy and calculate an associated mass using

$$m^2_{\text{recoil}} = E^2_{\text{CM}} + m^2_{\text{Obs}} - 2E_{\text{CM}}E_{\text{Obs}}, \text{(CM=Center of Mass \& Obs=Observed).} \quad (2)$$

Conceptually, it is directly seen the Z events will have a higher recoil mass since they recoil off the Higgs which is more massive than the W recoil off of the W giving a separation of ~45 GeV/c^2. Interestingly, the Z recoil mass distribution peaks above the Higgs mass which is explained by imperfect measurements and undetected neutrinos which occur more frequently in

the Z's hadronization channels ($b\bar{b}$, $c\bar{c}$) leading to missing energy and a more massive recoil as seen in Figure 12.

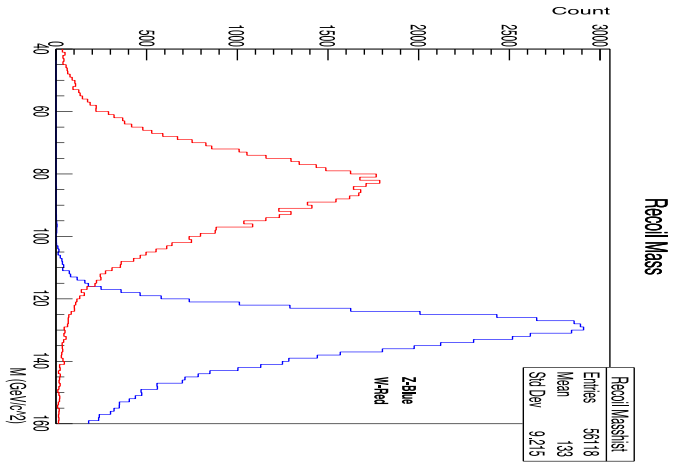

Figure 11: Comparison of Dijet Recoil Mass of W and Z Events

This plot shows the recoil mass of the combined dijet system using the PFO data. W is in red while Z is in blue. The plot displays an equal number of W and Z events.

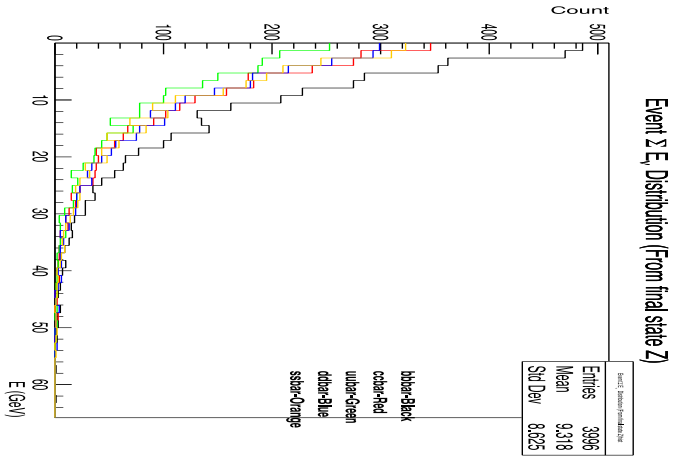

Figure 12: Distribution of Summed Neutrino Event Energy

This plot shows the summed energy of final state MCT neutrinos from Z events. Particularly, this distribution plots neutrinos from different quark flavors differently and shows $b\bar{b}$ event produce the most neutrinos. Importantly, these neutrinos are an important source of missing energy for the reconstructed energy of the PFO Z events since neutrinos are essentially invisible.

In addition to the recoil mass, the individual jet energies provide critical insight into the dynamics of these events which vary between the Z and W events. Importantly, W events are pair production events which means each boson will carry half of the collision energy which means the W dijets will have a combined energy of 125 GeV. On the other hand, the Higgs-strahlung process produces particles of varying masses meaning they will carry a different amount of energy. In this process, the kinematics demand the Z will carry 110 GeV while the Higgs will carry the rest. Even though the dijet systems will differ by around 15 GeV, the individual jets have the potential to differ more as seen in Figure 13a and Figure 13b. In this analysis, jet 1 will refer to the higher energy jet while jet 2 will be the lower energy jet giving us two additional parameters associated with the kinematics of these processes.

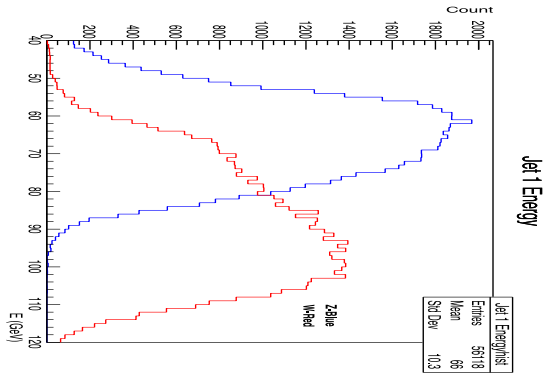

Figure 13a: Comparison of Energy of Jet 1 of W and Z Events

This plot shows the distribution of the energy of jet 1 from the PFO data. The W events have a mean around 88.6 GeV while Z events have a mean around 66 GeV. The plot displays an equal number of W and Z events.

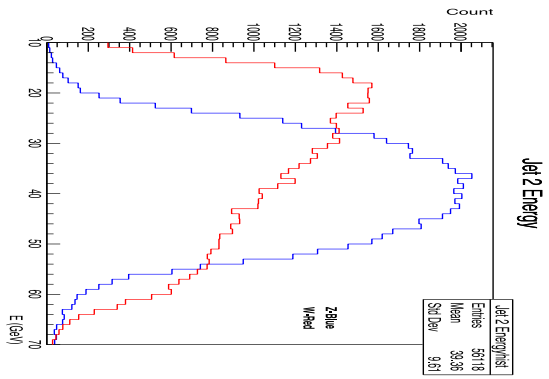

Figure 13b: Comparison of Energy of Jet 2 of W and Z Events

This plot shows the distribution of the energy of jet 2 from the PFO data. The W events have a mean around 33.8 GeV while Z events have a mean around 39.4 GeV. The plot displays an equal number of W and Z events.

Alongside kinematical parameters, we can utilize six unique parameters associated with the specific way the bosons interact with fermions. Particularly, the W boson exclusively couples to left-handed particles and right-handed anti-particles while the Z boson interacts slightly stronger to left-handed particles compared to right-handed particles. The difference in their interactions leads to varying properties in their production and decay processes. This analysis considers the eLpR scenario to maximize W production and identify separation parameters to suppress this overwhelming background as seen in Figure 14a and Figure 14b.

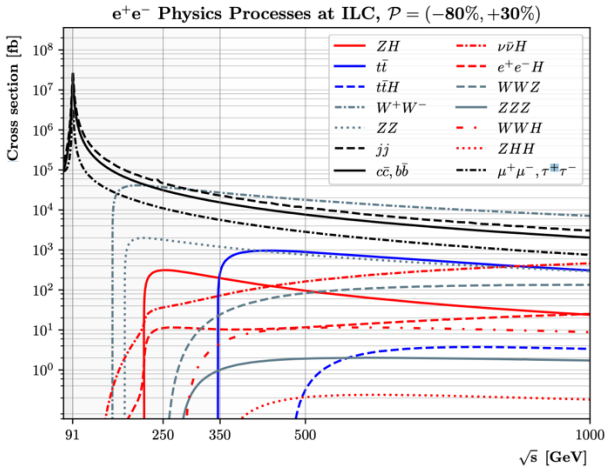

Figure 14a: Cross Section of SM Processes at the ILC with eLpR Polarization

This distribution shows the cross section of the critical SM processes at the ILC with the eLpR polarization as a function of center of mass energy. Importantly, the Higgs-strahlung process is vastly dominated by W pair production and Z pair production [3].

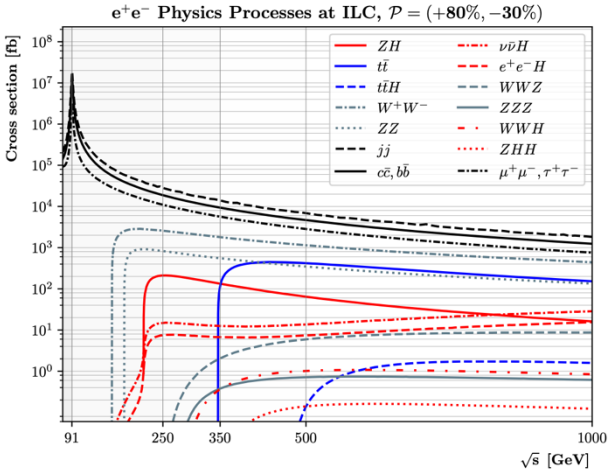

Figure 14b: Cross Section of SM Processes at the ILC with eRpL Polarization

This distribution is identical to Figure 14a except the cross section is calculated using the eRpL polarization. Critically, the domination of W and Z pair production in this distribution is not as overwhelming as Figure 14a [3].

The W^+ couples strongly to a right handed positron while the W^- couples strongly to a left handed electron. The eLpR polarization maximizes W pair production which makes it a dominating background. Importantly, the bosons have a strong likeliness to be produced in the direction of the particle with which they strongly interact. To create a sense of direction, the

incoming electron-positron beams can be thought of as forming the z-axis for measurements and the analysis. The combined dijet four vectors represent the reconstructed boson's four vector which can provide critical information about the boson's production. Particularly, the cosine of the angle of the boson's momentum and the beam axis (θ_B) can be used since the W^+ will most likely be along the direction of the positron which is the backwards direction ($\cos(\theta_B) \approx -1$). On the other hand, the Z will be generally randomly oriented as seen in Figure 15. The parameter can be directly calculated using the reconstructed momentum three vector using

$$\cos(\theta_B) = \frac{p_z}{|\vec{p}|} \tag{3}$$

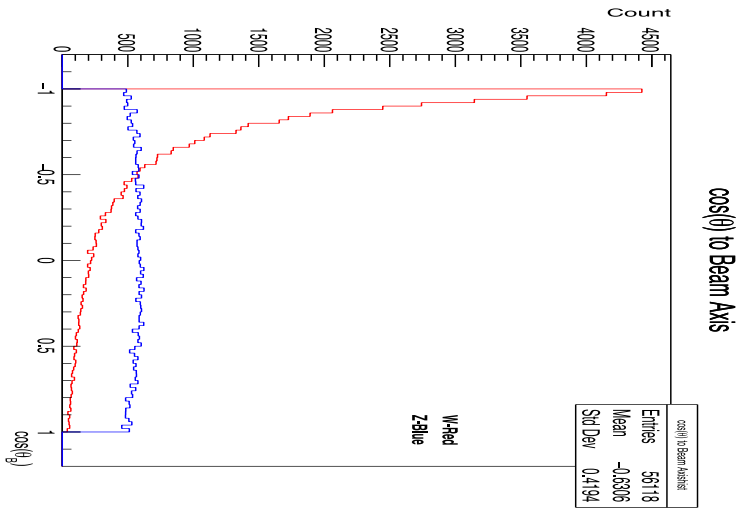

Figure 15: Comparison of Dijet Beam Angle of W and Z Events

This plot shows the distribution of the $\cos(\theta_B)$ of the PFO bosons reconstructed using the dijet system. Importantly, the maximal parity violating nature of the W is made evident by its sharp peak at -1 while the Z plot is randomly distributed. The plot displays an equal number of W and Z events.

In addition to the production of the bosons, their decays into quarks and eventually jets will differ based on their unique couplings. To preserve angular momentum, one of the quarks will be left-handed while the other one will be right-handed. Furthermore, the Z boson's interactions with the quarks can be understood using the non-flavor changing neutral weak

current so it will decay into a quark and an anti-quark of the same flavor. Particularly, the Z decays into down type quarks ($d\bar{d}$, $s\bar{s}$, $b\bar{b}$) around 15.2% of the time each while it decays into up type quarks ($u\bar{u}$, $c\bar{c}$, and $t\bar{t}$ is forbidden due to energy conservation) around 12% of the time each [10]. On the other hand, the W boson interactions are flavor changing for quarks so the interaction of the boson to them is understood through the elements of a unitary matrix referred to as the CKM matrix meaning the quark pairs produced will be of different flavors, and potentially different generation. Particularly, the W^+ decays into $u\bar{d}$ and $c\bar{s}$ around 32% of the time each, into $u\bar{s}$ and $c\bar{d}$ around 2% of the time each, into $c\bar{b}$ around 0.06% of the time, and into $u\bar{b}$ around 0.0005% of the time [10].

It can clearly be seen that the branching ratio of the W boson to b quarks is heavily constrained compared to the Z branching ratio. Importantly, a commonly used tool within particle physics to identify the flavor of a jet's originating quark are jet flavor tagging algorithms which utilize various tactics to see if it is likely that a jet originated from a b quark (and sometimes c quarks). In the past, these algorithms utilized single tagging (one jet) or double tagging (both jets) to measure the branching ratio of these bosons. Importantly, these quarks are the heavier quarks (excluding the top quark due to mass constraints) meaning they will travel a larger distance (the c less so than the b quark) compared to the lighter mass flavors and produce B and D mesons which aid in identifying the jet's flavor. Previous experiments and analyses like the SLD have achieved a b-tagging efficiency of 61.8% and a c-tagging efficiency of around 24.3% which can be modeled for a sense of realism [21]. Additionally, these flavor tagging models are not perfect so they will occasionally mistag jets. Particularly, the SLD study found that c jets and light flavor jets were tagged as b jets around 1.2% and 0.1% of the time

respectively. On the other hand, b jets and light flavor jets were tagged as c jets around 5.4% and 0.2% of the time respectively.

Specifically, we used MCT information where the higher energy quark was associated with jet 1 and the other one was associated with jet 2. Since the MC simulation provides particle type information, we were able to see exactly which jets originated from b and c quarks. To replicate tagging and mis-tagging efficiency, a random number was generated between 1 and 1000 for every jet, and the original quark flavor was used. Given the original quark is a bottom, the jet will be tagged as a charm given its random number is between 1 and 54 while a number between 54 and 672 means the jet will be tagged as a bottom. On the other hand, if the original quark is a charm, the jet will be tagged as a charm given its number is between 1 and 243 and it will be a bottom if its number is between 243 and 297. Finally, if the original quark is a light flavor, a jet with a number between 1 and 2 will be tagged as a charm, and a number between 2 and 3 will be tagged as a bottom.

Critically, we can construct four separation parameters from this information which will be the number of b quarks in jet 1 (b_1), the number of c quarks in jet 1 (c_1), the number of b quarks in jet 2 (b_2), and the number of c quarks in jet 2 (c_2). In the case of perfect efficiency, we expect (b_1, b_2, c_1, c_2) = (1, 1, 0, 0) or (0, 0, 1, 1) for the Z events which are cleanly separated from the W events which only produce events with two heavy quarks extremely rarely, and even then they will be of different flavors. However, the modeled efficiency does increase confusion making the separation more difficult, but it increases the realism of the analysis. Regardless, these separation parameters are still useful in the events as seen in Figure 16a-d.

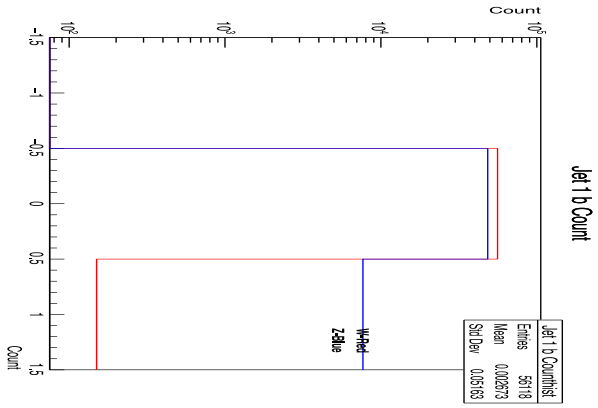

Figure 16a: Comparison of Jet 1 b Count of W and Z Events

This semi-log plot shows the distribution of the number of b jets tagged for jet 1 using the modeled efficiency for W and Z events. The plot displays an equal number of W and Z events.

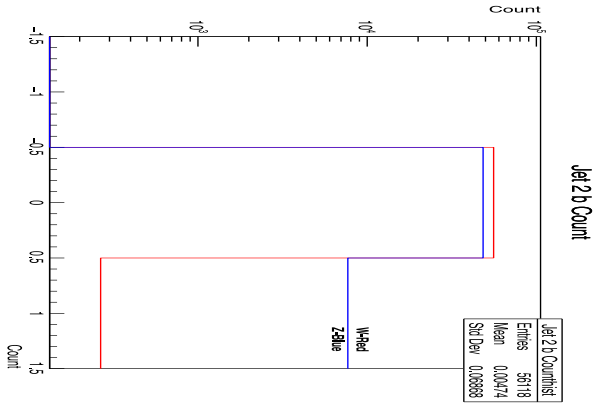

Figure 16b: Comparison of Jet 2 b Count of W and Z Events

This semi-log plot shows the distribution of the number of b jets tagged for jet 2 using the modeled efficiency for W and Z events. The plot displays an equal number of W and Z events.

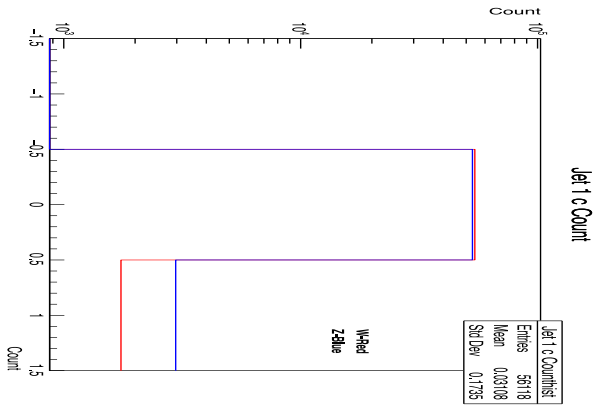

Figure 16c: Comparison of Jet 1 c Count of W and Z Events

This semi-log plot shows the distribution of the number of c jets tagged for jet 1 using the modeled efficiency for W and Z events. The plot displays an equal number of W and Z events.

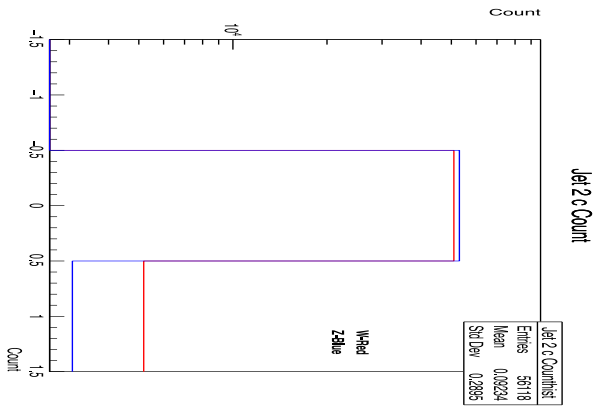

Figure 16d: Comparison of Jet 2 c Count of W and Z Events

This semi-log plot shows the distribution of the number of c jets tagged for jet 1 using the modeled efficiency for W and Z events. The plot displays an equal number of W and Z events.

Alongside the flavor of quarks produced by the boson, the handedness and varying mass of the quarks produced can have a useful kinematical effect in these events. Specifically, the right-handed anti-quark will be aligned with the W boson meaning its momentum three vector will be some deviation off the -z direction. Moreover, the W^+ decays imply this anti-quark will be the heavier quark. Given the other quark is lighter and its momentum vector will be some deviation off the xy plane, we can see the opening angle (θ_O) between these quarks and the eventual jets will be around 90° ($\cos(\theta_O) \approx 0$). On the other hand, the quarks decaying from the Z will not have as strong of a preference for their alignment with the boson and the pair will have an equal mass which means their opening angle will be larger compared to the W which can be seen in Figure 17.

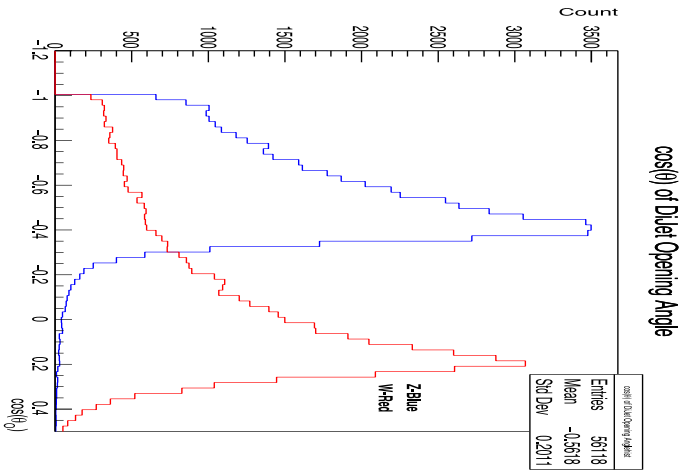

Figure 17: Comparison of Dijet Opening Angle of W and Z Events

This plot shows the distribution of the $\cos(\theta_O)$ of the PFO dijets. Importantly, the plots peaks in different regions namely above and below 0. The plot displays an equal number of W and Z events.

Before training a neural network to separate the signal Z events from the background W events, these powerful separation parameters can be used to clean the data to focus on events that are not obviously background. The goal of these four strict cuts will be to minimally remove Z events to maintain a high signal efficiency while removing substantial W events to maximize background rejection. The first cut will be on the mass to remove light W events which have a smaller likelihood of being a Z. Specifically, we only focus on combined dijets with an invariant mass between 80 GeV/c^2 and 105 GeV/c^2 where ~88% of our Z events survive and ~45% of our W events survive. Next, we will cut on the recoil mass to focus on events where the dijet system recoiled off a Higgs, so we require the recoil mass to be between 100 GeV/c^2 and 160 GeV/c^2 where around 88% of our Z events still survive but only ~2% of our W events survive. Thirdly, we implement a cut on $\cos(\theta_B)$ since the W^+ bosons are primarily in the backward direction so we will require our dijet system to have $\cos(\theta_B) \geq -0.6$. This leads to ~71% of our Z events survive and ~1% of our W events survive. Finally, we implement a cut on $\cos(\theta_O)$ to focus on

particles with a larger opening angle so we require our dijet system to have $\cos(\theta_O) \leq -0.2$ which leads to ~60% of our Z events surviving and ~0.24% of our W events surviving. Figures 18a-j shows the distributions of the various separation parameters following these cuts.

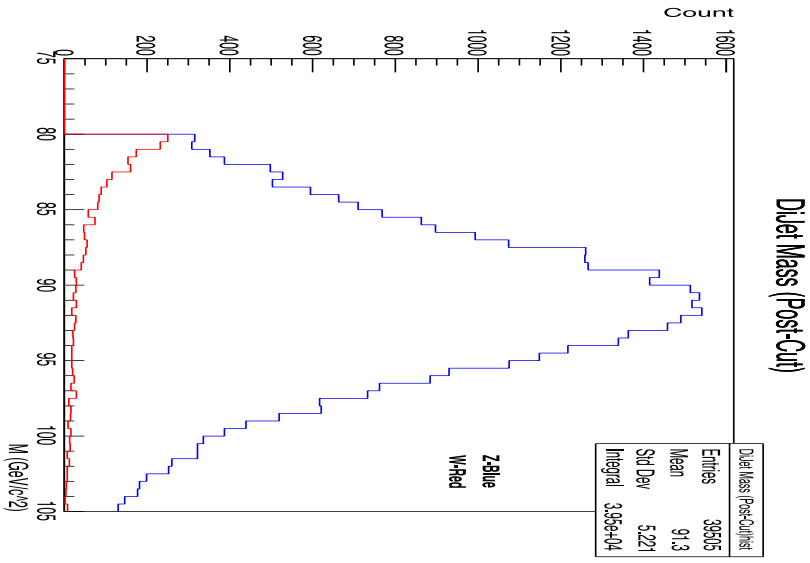

Figure 18a: Post-Cut Comparison of Dijet Invariant Mass of W and Z Events

This plot displays the dijet invariant mass distribution following our 4 data cuts.

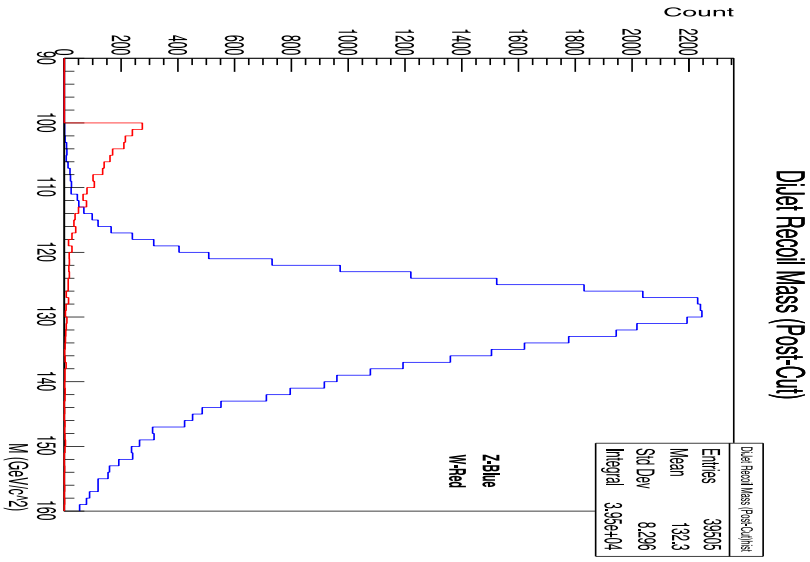

Figure 18b: Post-Cut Comparison of Dijet Recoil Mass of W and Z Events

This plot displays the dijet recoil mass distribution following our 4 data cuts.

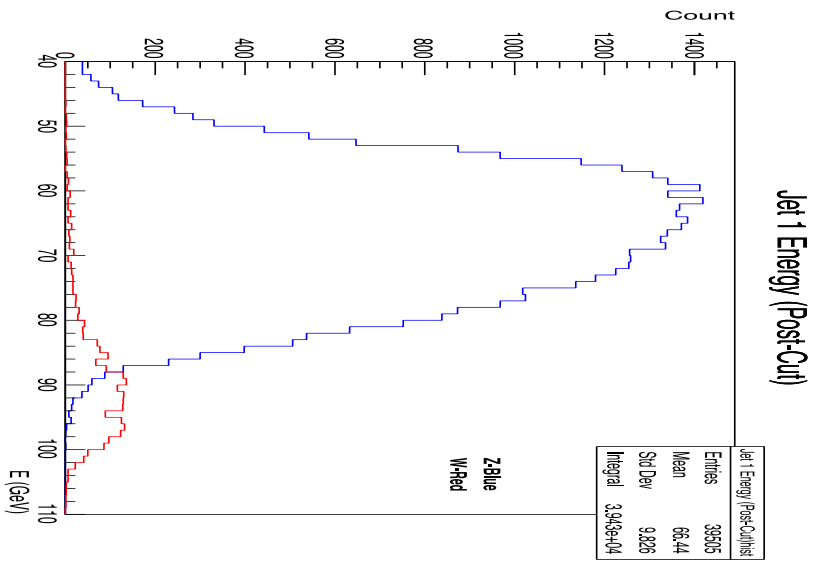

Figure 18c: Post-Cut Comparison of Jet 1 Energy of W and Z Events

This plot displays the energy of jet 1 following our 4 data cuts.

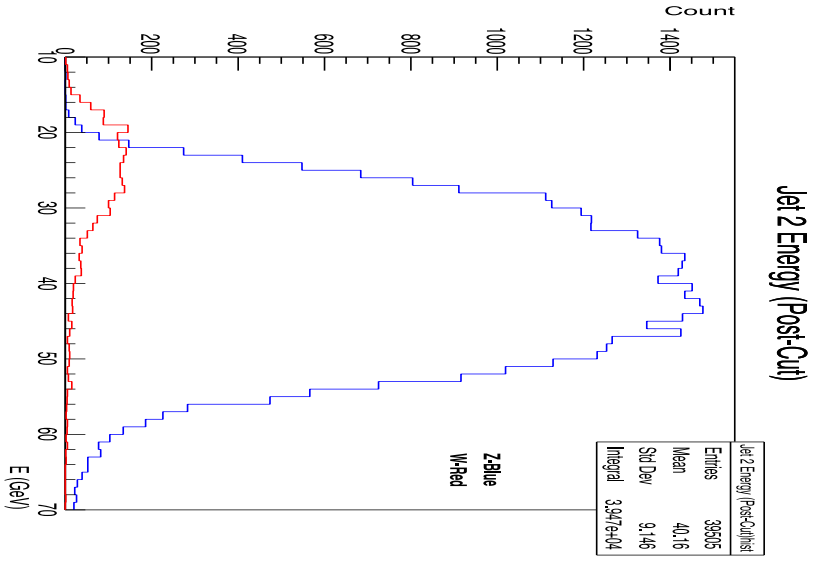

Figure 18d: Post-Cut Comparison of Jet 2 Energy of W and Z Events

This plot displays the energy of jet 2 following our 4 data cuts.

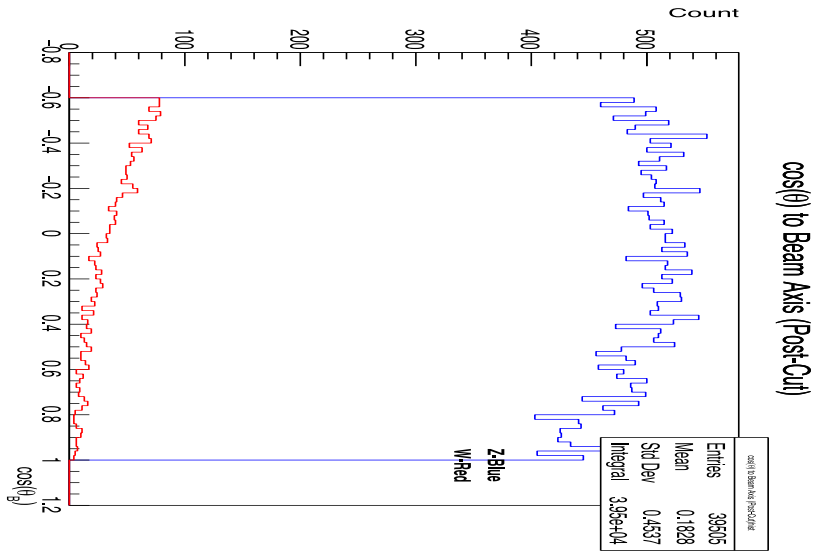

Figure 18e: Post-Cut Comparison of Dijet Beam Angle of W and Z Events

This plot displays the $\cos(\theta_B)$ following our 4 data cuts.

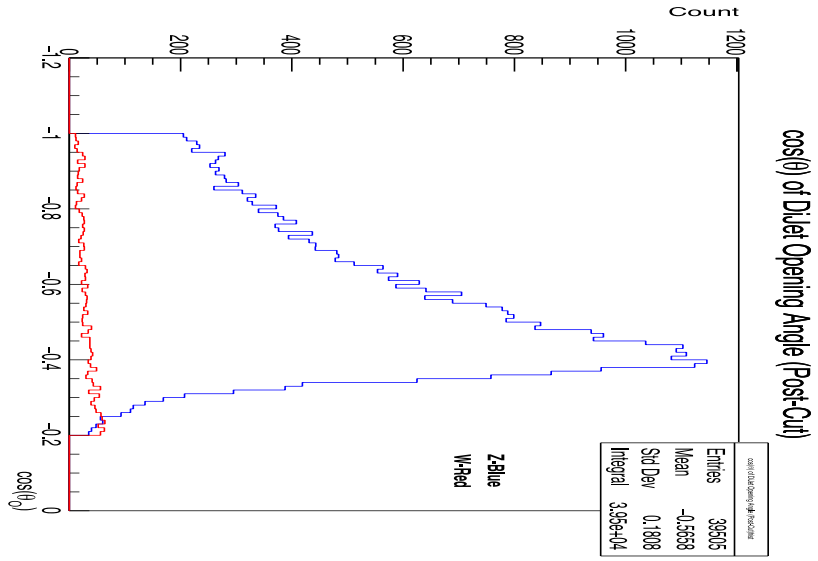

Figure 18f: Post-Cut Comparison of Dijet Opening Angle of W and Z Events

This plot displays the $\cos(\theta_O)$ following our 4 data cuts.

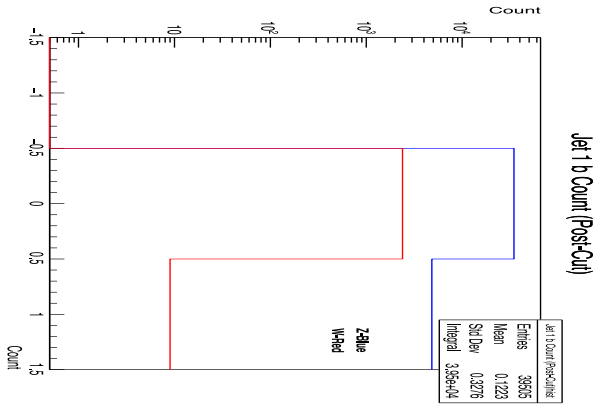

Figure 18g: Post-Cut Comparison of Jet 1 b Count of W and Z Events

This semi-log plot displays b_1 following our 4 data cuts.

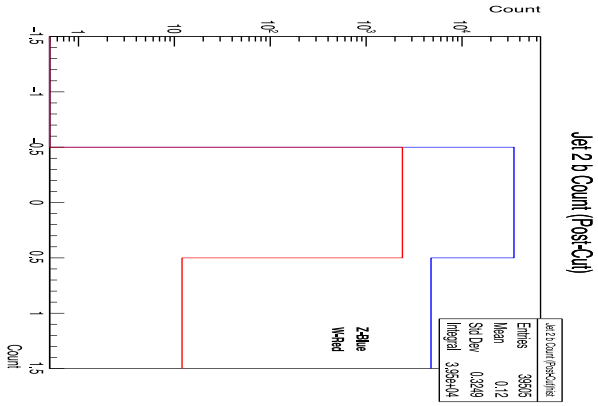

Figure 18h: Post-Cut Comparison of Jet 2 b Count of W and Z Events

This semi-log plot displays b_2 following our 4 data cuts.

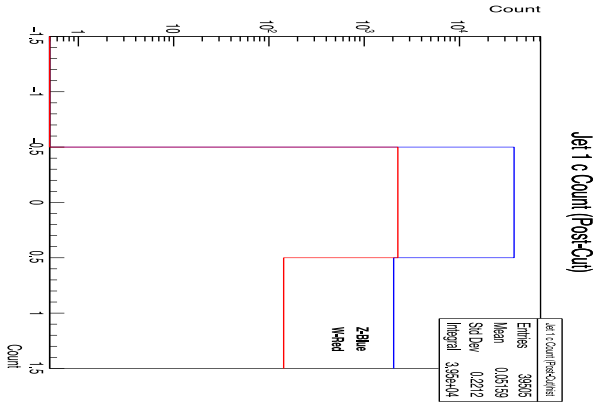

Figure 18i: Post-Cut Comparison of Jet 1 c Count of W and Z Events

This semi-log plot displays c_1 following our 4 data cuts.

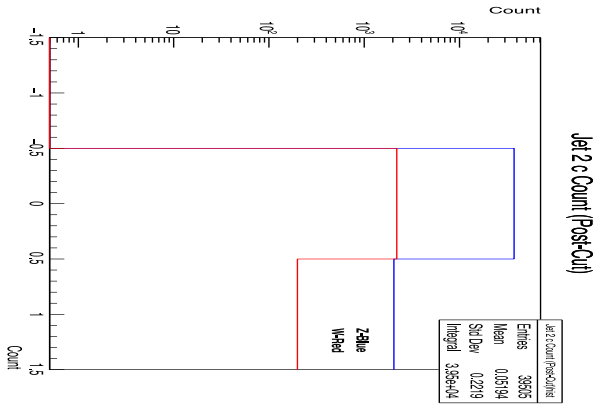

Figure 18j: Post-Cut Comparison of Jet 2 c Count of W and Z Events

This semi-log plot displays c_2 following our 4 data cuts.

Training and Evaluating a Separator

Following the incredibly effective cuts using the separation parameters, we utilized a neural network to classify a given event as a signal event (Z) or a background event (W) using the 10 separation parameters we developed. Specifically, we used the built-in toolkit for multivariate analysis in ROOT referred to as TMVA [22][23]. Typically, a neural network will take in various variables and pass them through one or two layers of nodes to give a desired output. In this model, a node will weigh values from the previous layer and combine them to pass onto the next layer. Weights connecting nodes are determined during the training stage of the neural network. Once a neural network is trained on a training dataset, the effectiveness of the training is evaluated using the model on the training data. To evaluate the effectiveness of the model, the neural network is then utilized on additional unbiased data.

Since this analysis was performed in Python, PyKeras was used as the neural network model which allows for Python implementation of TensorFlow and Keras. Particularly, we used a deep neural network for this separation analysis which is a neural network with more than two layers of nodes between the input and output to separate the signal from the background. Crucially, the output for every event is in the form of a number between 0 and 1 associated with strong confidence of being a background and signal event respectively. A cut can be taken on this neural network weight to identify what the model calls signal.

This analysis reduced every dijet event into the ten separation parameters which were put through rigorous cuts to minimize the number of background events. Following the cuts, we were left with 2400 W events and 39505 Z events to train on and test. We chose 1440 W events and 6000 Z events to train our neural network. Since the model will not be a perfect separator,

any given cut on the neural network weight will result in a specific tradeoff between the signal efficiency and background rejection which can be seen in the ROC curve in Figure 19.

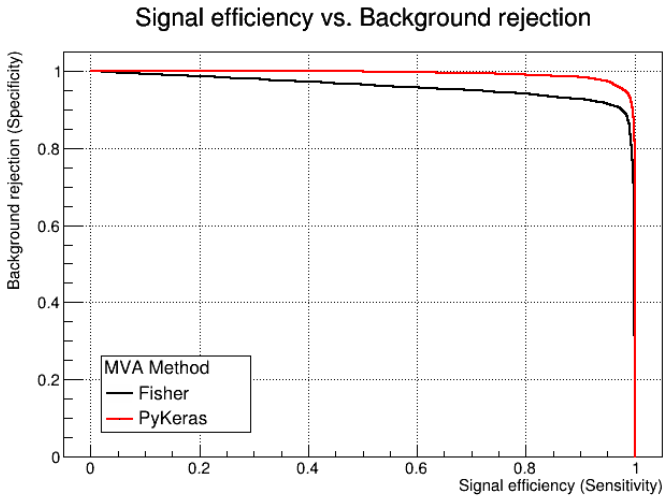

Figure 19: ROC Curve for Training a W and Z Classifier

This plot shows the respective background rejection for a given signal efficiency which gives a method of evaluating the training's effectiveness.

Importantly, the scatter plots like in Figure 20a-j show the neural network weight vs the value of a particular separation parameter of each event to illustrate the significance the model placed on the various parameters. Critically, the scatter plots for the b count parameters show events with a tagged b jet provide the model with strong confidence the event is a signal event as seen in Figure 20g and 20h. Additionally, the recoil mass scatter plot shows a heavy recoil mass is indicative of the model being more confident the event is a signal event as seen in Figure 20b. Moreover, the other scatter plots collectively show the model is accurately separating the events based on the separation parameters as intended.

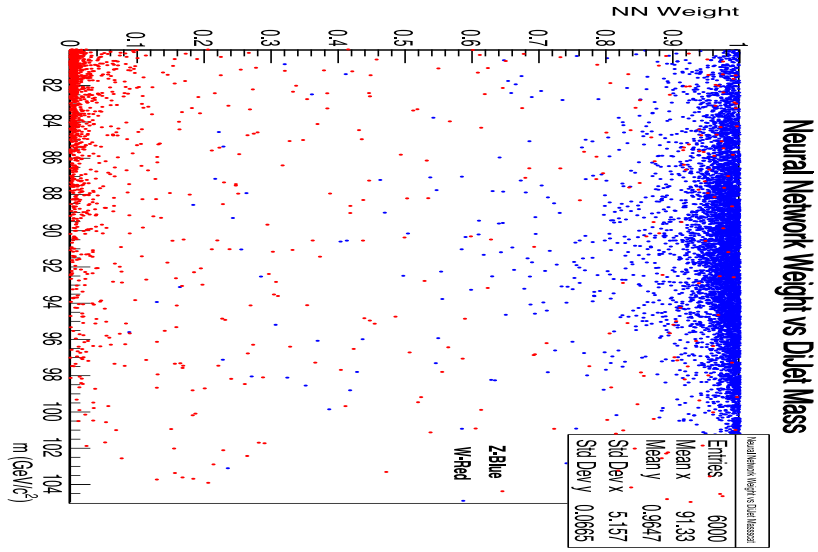

Figure 20a: Impact of W and Z Dijet Invariant Mass on Training

This scatter plot shows how the neural network training evaluated the dijet invariant mass value of each event.

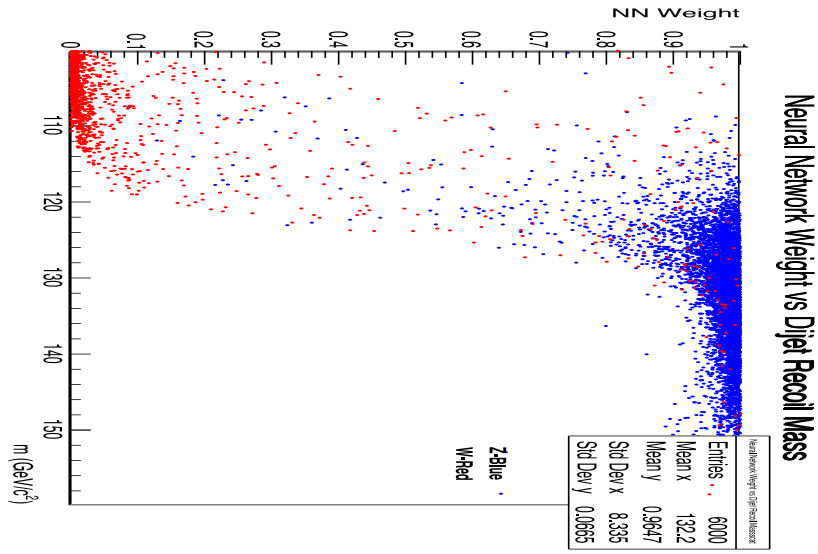

Figure 20b: Impact of W and Z Dijet Recoil Mass on Training

This scatter plot shows how the neural network training evaluated the recoil mass of each event.

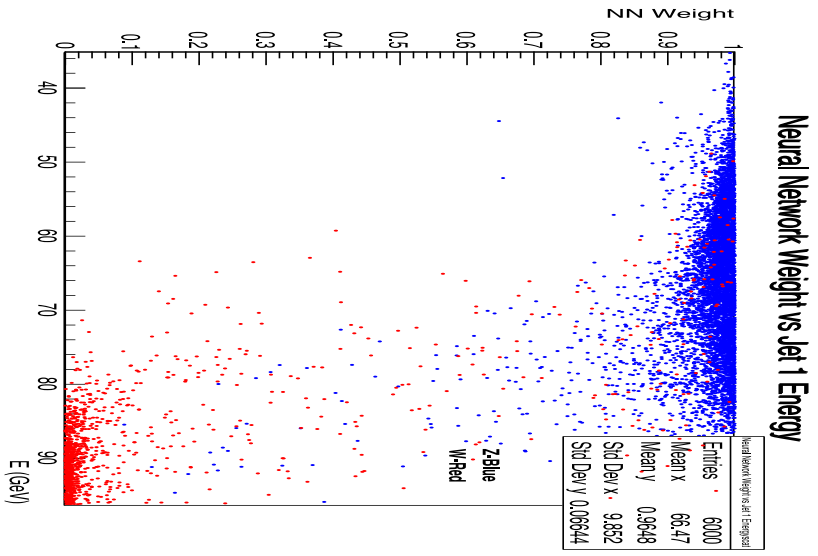

Figure 20c: Impact of Jet 1 Energy of W and Z Events on Training

This scatter plot shows how the neural network training evaluated jet 1 energy of each event.

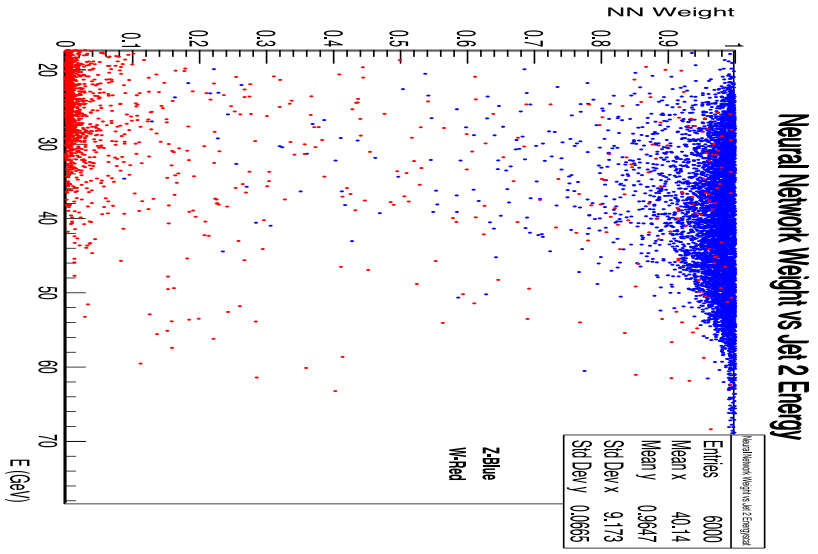

Figure 20d: Impact of Jet 2 Energy of W and Z Events on Training

This scatter plot shows how the neural network training evaluated jet 2 energy of each event.

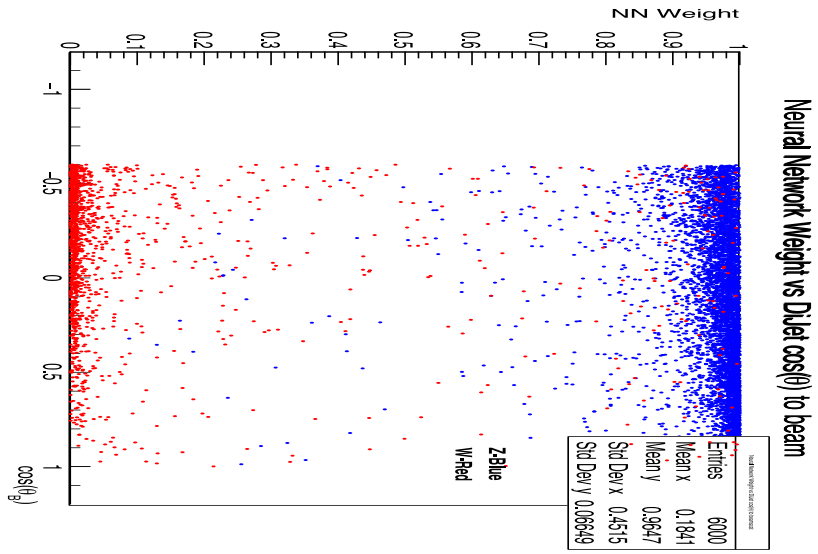

Figure 20e: Impact of W and Z Dijet Beam Angle on Training

This scatter plot shows how the neural network training evaluated the $\cos(\theta_B)$ of each event.

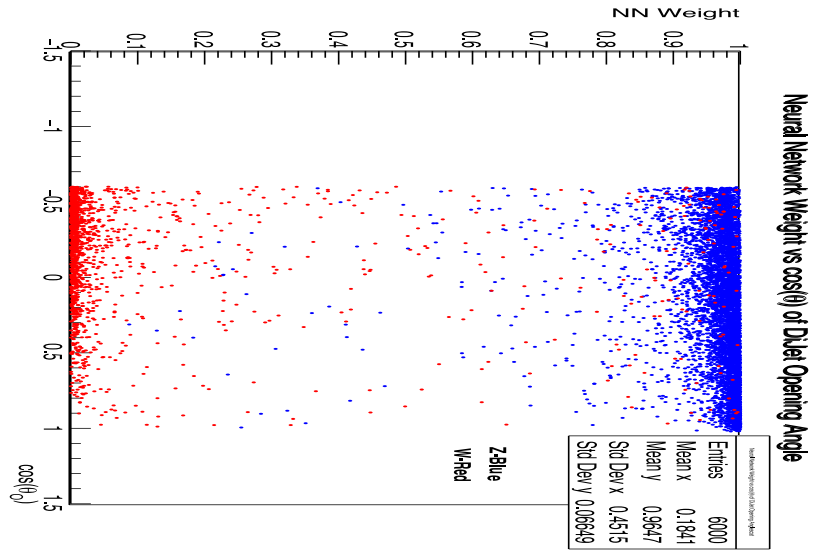

Figure 20f: Impact of W and Z Dijet Opening Angle on Training

This scatter plot shows how the neural network training evaluated the $\cos(\theta_O)$ of each event.

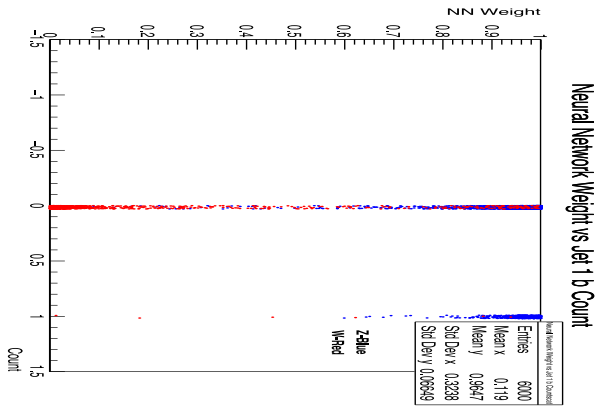

Figure 20g: Impact of Jet 1 b Count of W and Z Events on Training

This scatter plot shows how the neural network training evaluated b_1 of each event.

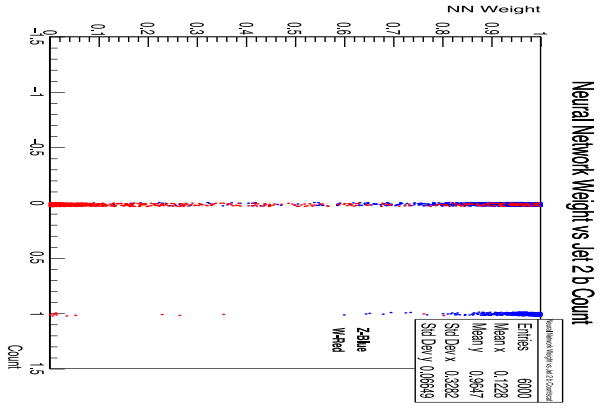

Figure 20h: Impact of Jet 2 b Count of W and Z Events on Training

This scatter plot shows how the neural network training evaluated b_2 of each event.

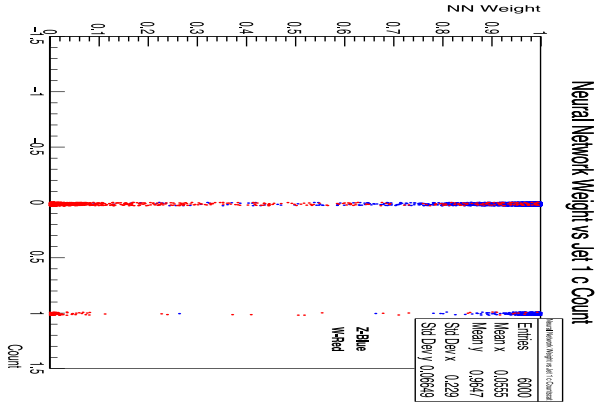

Figure 20i: Impact of Jet 1 c Count of W and Z Events on Training

This scatter plot shows how the neural network training evaluated c_1 of each event.

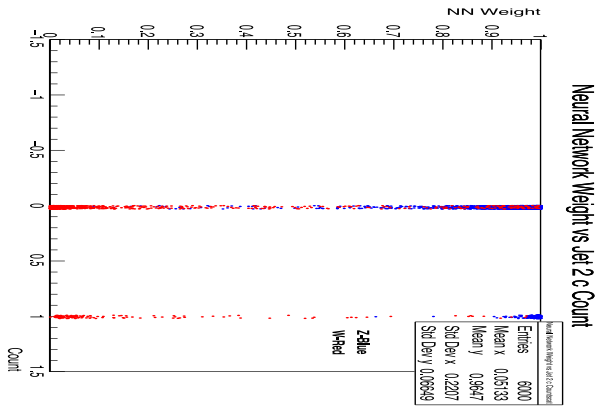

Figure 20j: Impact of Jet 2 c Count of W and Z Events on Training

This scatter plot shows how the neural network training evaluated c_2 of each event.

After training the neural network, the model can be used on the remaining 960 W events and 4008 Z events to evaluate the effectiveness of the classifier in separating signal and background. Importantly, the goal is to choose a neural net weight that maximizes the number of signal events and minimizes the number of background events using Figures 21a and 21b. Requiring events to have a neural net weight above 0.97 gives a signal efficiency of 69.2% and a background rejection of 97.3% which results in 2798 signal events where 26 are W events. The spread and influence of these W events within the distribution of the signal events in the various separation parameters can be seen in Figures 22a-j.

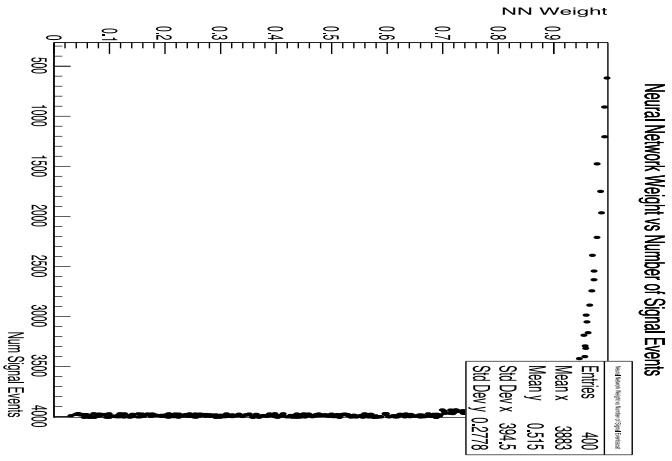

Figure 21a: Neural Network Weight and Signal Efficiency

This plot shows the required neural network weight cut needed for a desired number of signal events.

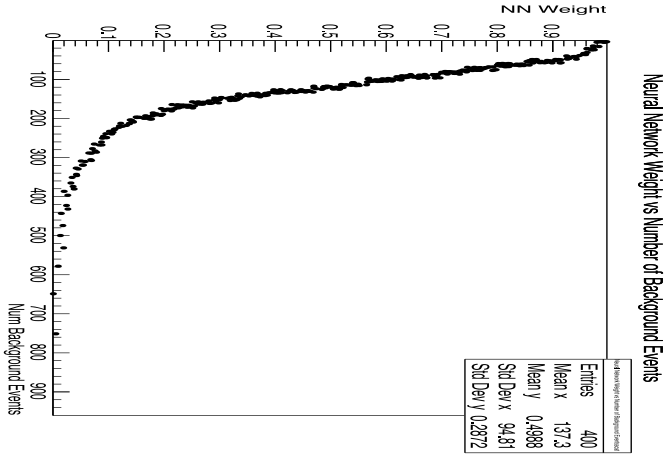

Figure 21b: Neural Network Weight and Signal Efficiency

This plot shows the required neural network weight cut needed for a desired number of signal events.

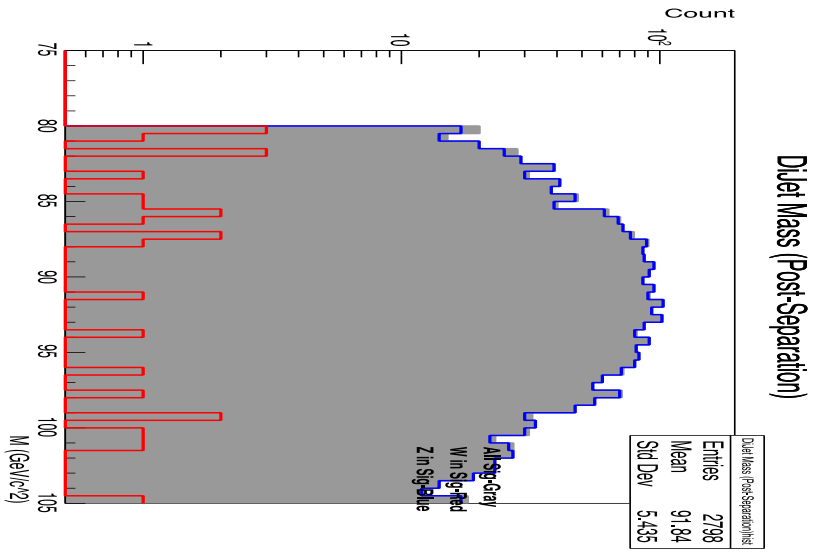

Figure 22a: Post-Separation Signal Mass Distribution

This semi-log plot shows the mass distribution of the signal broken up into the contributions from Z and W events.

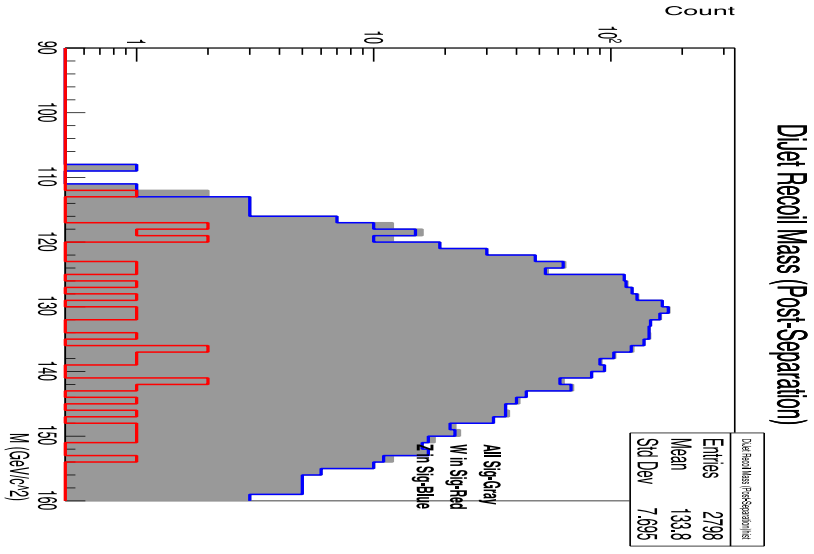

Figure 22b: Post-Separation Signal Recoil Mass Distribution

This semi-log plot shows the recoil mass distribution of the signal broken up into the contributions from Z and W events.

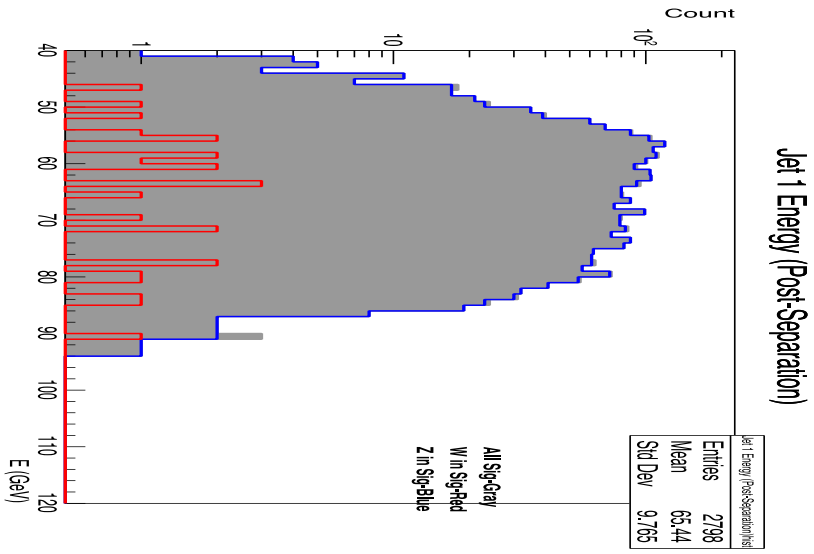

Figure 22c: Post-Separation Signal Jet 1 Energy Distribution

This semi-log plot shows the jet 1 energy distribution of the signal broken up into the contributions from Z and W events.

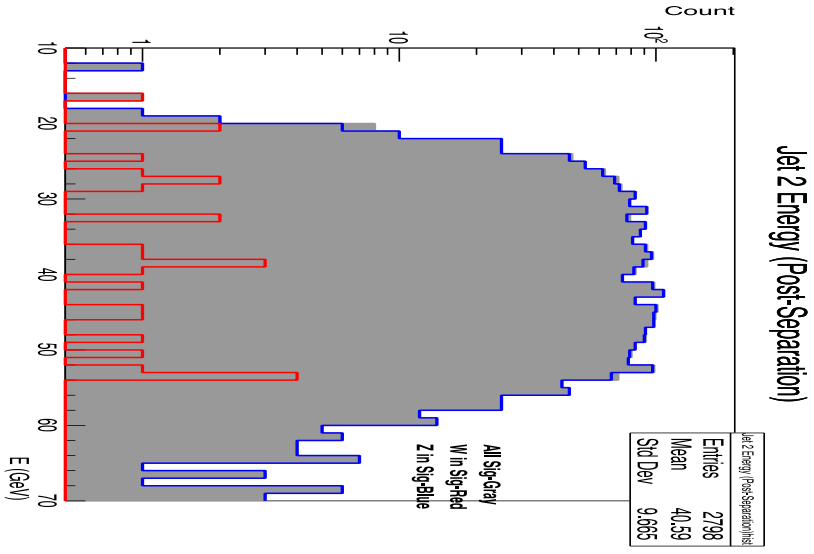

Figure 22d: Post-Separation Signal Jet 2 Energy Distribution

This semi-log plot shows the jet 1 energy distribution of the signal broken up into the contributions from Z and W events.

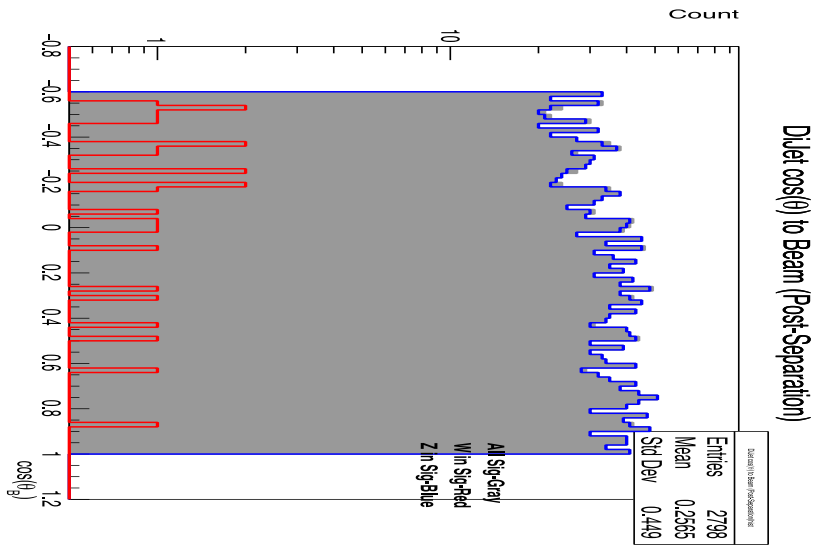

Figure 22e: Post-Separation Signal Dijet Beam Angle Distribution

This semi-log plot shows the $\cos(\theta_B)$ distribution of the signal broken up into the contributions from Z and W events.

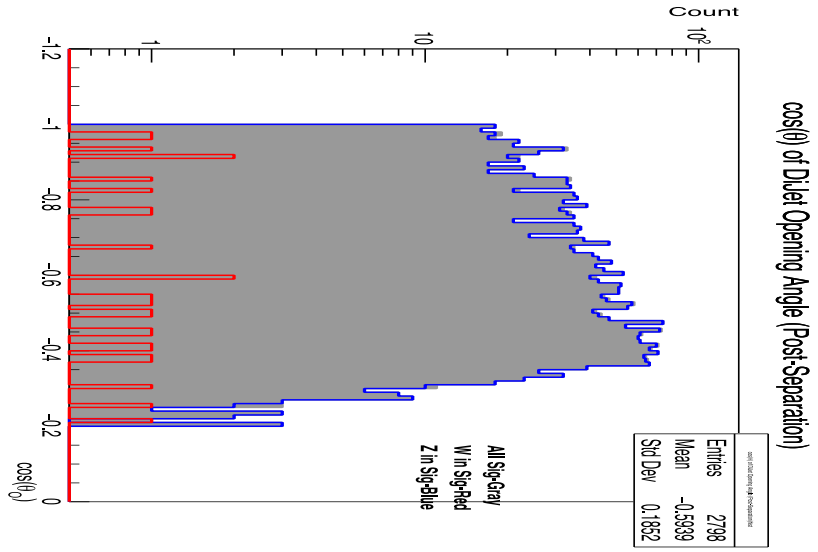

Figure 22f: Post-Separation Signal Dijet Opening Angle Distribution

This semi-log plot shows the $\cos(\theta_O)$ distribution of the signal broken up into the contributions from Z and W events.

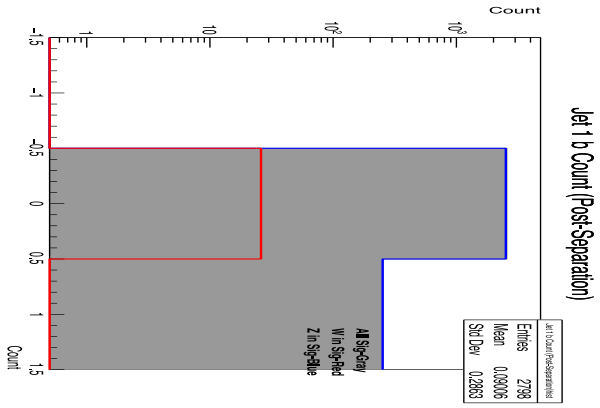

Figure 22g: Post-Separation Signal Jet 1 b Count Distribution

This semi-log plot shows the b_1 distribution of the signal broken up into the contributions from Z and W events.

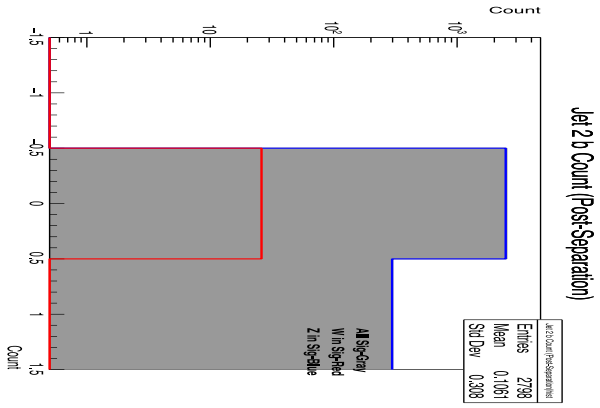

Figure 22h: Post-Separation Signal Jet 2 b Count Distribution

This semi-log plot shows the b_2 distribution of the signal broken up into the contributions from Z and W events.

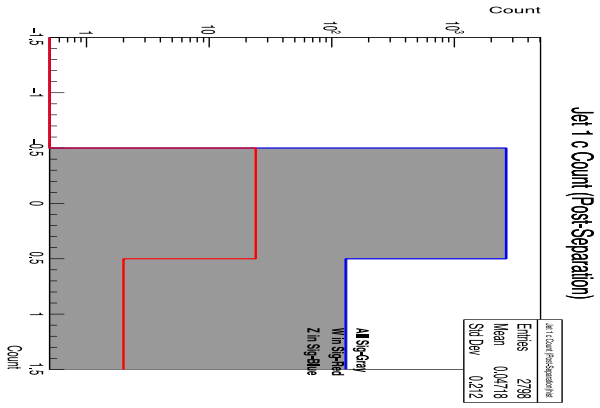

Figure 22i: Post-Separation Signal Jet 1 c Count Distribution

This semi-log plot shows the c_1 distribution of the signal broken up into the contributions from Z and W events.

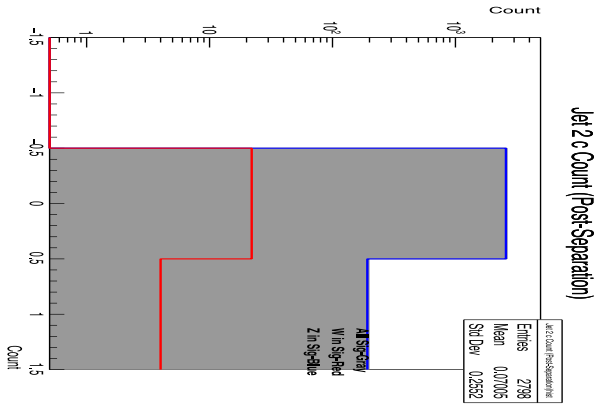

Figure 22j: Post- Separation Signal Jet 2 c Count Distribution

This semi-log plot shows the c_2 distribution of the signal broken up into the contributions from Z and W events.

Post-Separation Analysis

The fact our trained model leads to a signal composed of ~1% of W events and a 97.3% background rejection might seem overwhelmingly successful; however, it is critical to note these few background events are not weighed correctly since these W pair production events and Higgs-strahlung events do not occur at an equal rate. Particularly, the number of W events to the number of Z events is given by

$$N_{\frac{W}{Z}} = \frac{\sigma(e^+e^- \to W^+W^-)\ \Gamma\left(W^+ \to q_i\bar{q}_j\right)\ \Gamma\left(W^- \to \mu^-\bar{\nu}_\mu\right)}{\sigma(e^+e^- \to ZH)\ \Gamma(Z \to q\bar{q})\ \Gamma(H \to \text{invisible})} \tag{4}$$

This quantity can be easily evaluated by noting the cross section for W pair production at 250 GeV with beamstrahlung and beam polarization of eLpR gives is given by 37.9 nb and 0.297 nb for Higgs-strahlung. Noting the branching ratio of W into muonic leptons is 10.63%, we get 12485 W events for every Z event. In conjunction with the cut and dijet efficiencies, we see that every W event will need to be weighed by ~208. Importantly, the increased weight of W events would overwhelm the signal meaning we will need to reduce the signal efficiency by choosing a stricter neural net weight or develop a post-separation method to remove the W events.

This analysis primarily focused on the dijets themselves as the basis for separation since there are instances where the recoil off the W^+ can be missed entirely leading to a seemingly invisible recoil or where the Higgs decays leptonically where only a partial observation can be made which could replicate the W^+ recoil. Regardless, we can take advantage of the recoils of the dijets to further enrich our analysis and improve our separation. Specifically, every signal event can be analyzed to find the highest energy lepton within the event and the cosine of the angle between this lepton's momentum and the combined dijet system's momentum (θ_L).

We expect these parameters to be powerful in removing the W events from the signal since the highest energy lepton in the W events will be the lepton from the W^- which will have

more energy than any lepton produced in the jets of the Z boson. Moreover, the angle θ_L provides critical insight since the lepton produced in the jet production stage of the Z will be closer to the dijet momentum ($\theta_L \approx 0$) while the lepton produced during the W^- decay will be much farther away from the dijet momentum ($\theta_L \approx \pi$). The removal of most of the W events from the signal can be easily seen by imposing a cut on the phase space of these two parameters as seen in Figure 23a. Importantly, we can see in Figure 23b that we can relax our neural net weight to improve signal efficiency since this post-classifier separation can be powerful in removing W events.

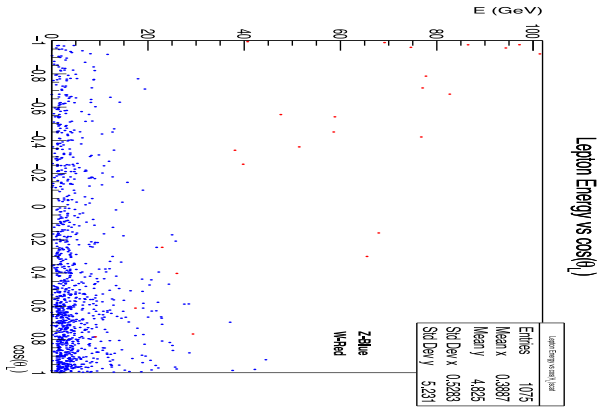

Figure 23a: Lepton Angle and Energy Phase Space for Signal ≥ 0.97

This scatter plot shows the phase space of the energy of the highest energy lepton in every event and the angle it forms with the combined dijet system. Cuts can be imposed to remove W events post-separation by trained neural network.

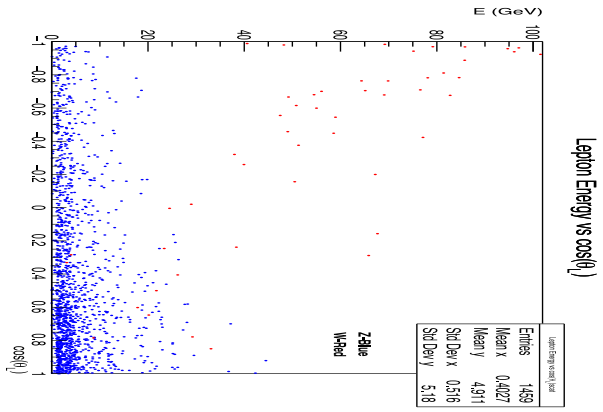

Figure 23b: Lepton Angle and Energy Phase Space for Signal ≥ 0.90

This scatter plot is the same plot as 23b with a relaxed neural net weight cut. With this relaxed cut of 0.9, we would have a signal efficiency of 94.2% and background rejection of 94.5% which leads to 53 W events remaining within the signal. Importantly, general cuts on this phase space will remove most of these events.

Regardless of the handful of W events remaining in the signal, the separation can be improved immensely by noting energy dependent separation parameters can be more discriminatory between signal and background with increased energy precision. As seen in Figure 24a and 24b, the reconstructed jet energy distributions align with the simulated distributions but there are clear areas to improve which can be seen using the scaled energy resolution defined as the standard deviation of the scaled energy difference distribution defined as

$$\frac{\Delta E}{\sqrt{E}} = \sqrt{E_{\mathrm{MCT}}}\left(1 - \frac{E_{\mathrm{PFO}}}{E_{\mathrm{MCT}}}\right). \tag{5}$$

The scaled jet energy resolution is $\frac{38.9\%}{\sqrt{E}}$ given by Figure 25 quantifies the inaccuracy of the detector in a manner that doesn't scale with energy.

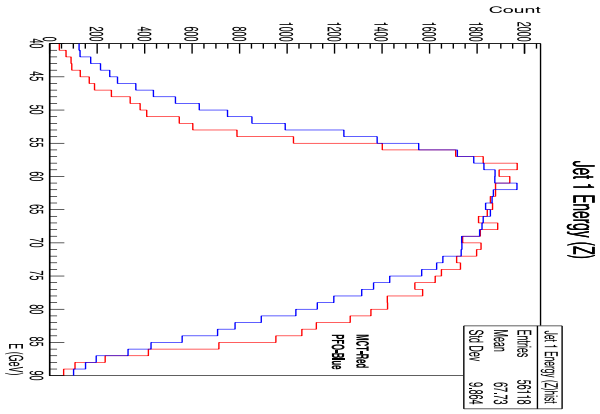

Figure 24a: Jet 1 Energy Distribution of Z Events

This distribution shows the energy distribution of jet 1 of the simulated dataset in red and reconstructed dataset in blue from Z events.

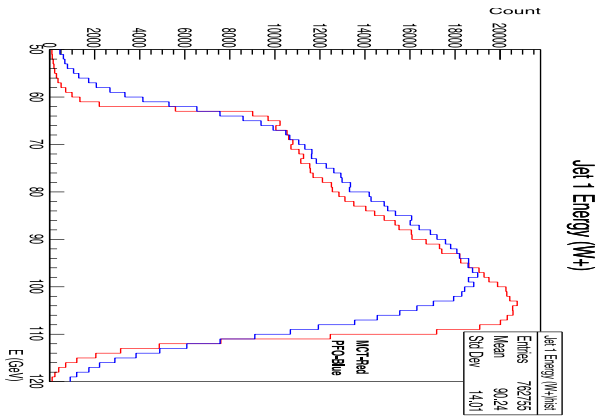

Figure 24b: Jet 1 Energy Distribution of W Events

This distribution shows the energy distribution of jet 1 of the simulated dataset in red and reconstructed dataset in blue from W events.

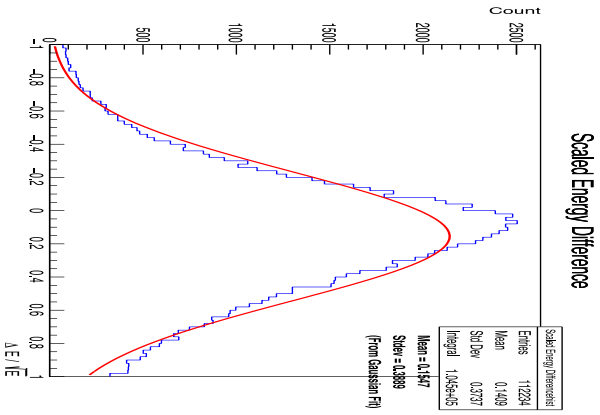

Figure 25: Scaled Jet Energy Resolution

This distribution shows the scaled energy difference of jets from Z events. The standard deviation of the Gaussian fit gives the scaled energy resolution.

Furthermore, we can note that a jet can be broken up into the various final state particle types that constitute including photons, leptons (e, μ), neutral hadrons (n, K_L^0), and charged hadrons (p, π, K). On average, 26% of the jet's energy is within photons, 1.9% of the energy is within leptons, 11% of the energy is within neutral hadrons, and 60.9% of the energy goes to charged hadrons as seen in Figures 26a-d. Since photons are entirely measured by the ECal, their contribution to the jet's energy is critical since any optimization to the calorimeter would improve the reconstruction of photons which inevitably improves the reconstruction of jets. As can be seen in Figure 27, the dependence of a jet's reconstruction accuracy increases with the increased contribution to its energy from photons. The improved resolution of the higher granularity MAPS model in conjunction with utilizing weighted clusters will improve the reconstruction of photons which will have a significant impact on the jets themselves. Not only would this improve the jet energy parameters, but the invariant mass and recoil mass parameters could improve drastically since it would reduce the width of the distribution increasing the separation of W and Z events.

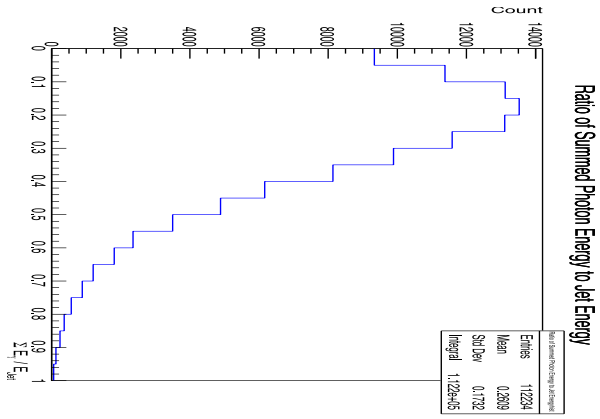

Figure 26a: Jet Energy in Photons

This histogram shows the ratio of how much of a jet's energy is in photons using the MCT Z data.

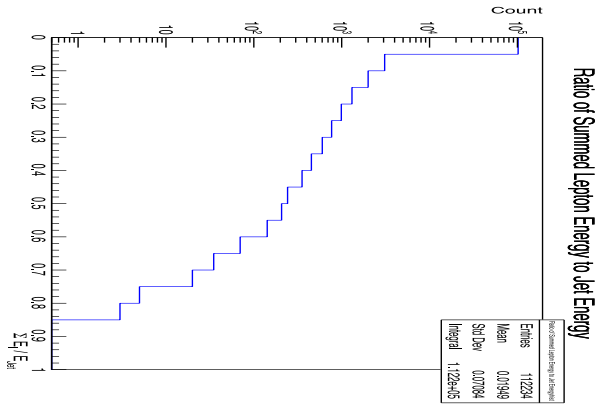

Figure 26b: Jet Energy in Leptons

This semi-log histogram shows the ratio of how much of a jet's energy is in leptons using the MCT Z data.

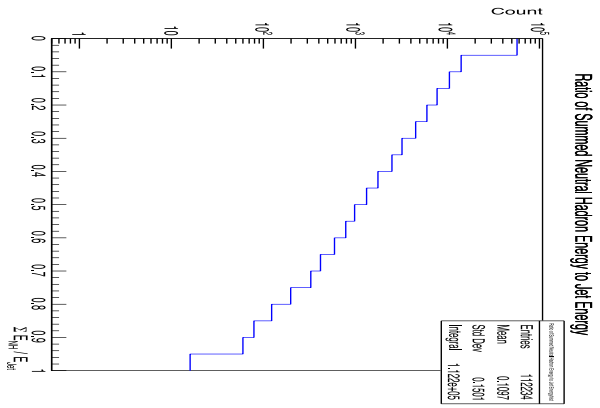

Figure 26c: Jet Energy in Neutral Hadrons

This semi-log histogram shows the ratio of how much of a jet's energy is in neutral hadrons using the MCT Z data.

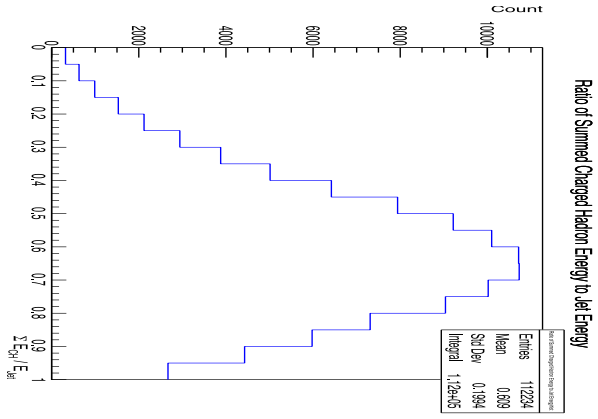

Figure 26d: Jet Energy in Charged Hadrons

This histogram shows the ratio of how much of a jet's energy is in charged hadrons using the MCT Z data.

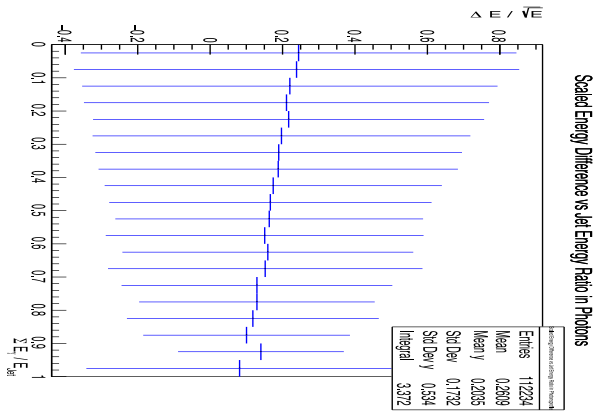

Figure 27: Scaled Jet Energy Resolution and Jet Energy in Photons

This profile plot graphs the percentage of a jet's energy in its constituent photons on the x-axis. Each bin has a horizontal line that corresponds to the mean scaled energy difference while the vertical line depicts the standard deviation which is the scaled energy resolution. Importantly, the vertical bars shorten as the jets have a greater concentration of photonic energy which implies the ECal is crucial for accurately measuring the jets since photons are exclusively detected within the ECal.

Conclusion

Following the utilization of the trained neural network, a general cut of $E_L \leq 15$ GeV can be imposed on the phase space of Figure 23b giving a final count of 3701 Z events and 6 W events as seen in Figure 28. Noting the correct weighing of W events, we arrive at the final Z to W event ratio of 3701: 1248 which is approximately 3 Z events for each W event. Even though this separation analysis has removed most of the W pair production background, the events still prove to be a large source of background. Specifically, every W point in Figure 28 would have 208 W events which would be randomly distributed about those points. Critically, the separation of Higgs-strahlung dijet events with invisible Higgs decay from the overwhelming background of W pair production dijet events can improved with increased energy precision. Particularly, the improved energy resolution of higher granularity calorimetry like the MAPS design can optimize the separability of the energy dependent separation parameters.

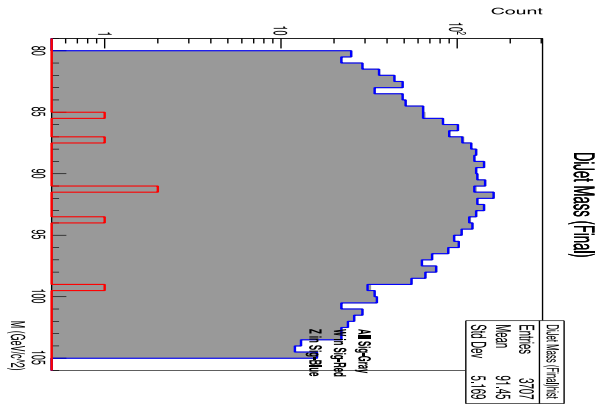

Figure 28: Invariant Dijet Mass Distribution Post Leptonic Energy Cut

This semi-log plot shows the distribution of the dijet invariant mass following the cut on events with leptons with energies higher than 15 GeV.

This study provided and explored critical physics-based parameters that can help remove one significant source of background from Higgs-strahlung events where the Higgs decays invisibly. This channel will be crucial to probing new physics BSM since the Higgs could decay

invisibly more frequently than expected. Importantly, the W pair production background is not the only large source of background since Z pair production events would also be substantial. A rich direction for future research includes improving and identifying more powerful parameters to further remove W pair production and Z pair production events. Additionally, this analysis relied upon the eLpR beam polarization in constructing separation parameters which might not be as useful when working with a different beam polarization or other more powerful parameters can be found for the other polarizations.

References

[1] G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." Phys. Let. B. 716, 1 (2012).

[2] T. Behnke et al., "The International Linear Collider Technical Design Report-Volume 1: Executive Summary." 2013. arXiv: 1306.6327.

[3] A. Aryshev et al., "The International Linear Collider: Report to Snowmass 2021." January 2023. arXiv: 2203.07622.

[4] T. Behnke et al., "The International Linear Collider Technical Design Report-Volume 4: Detectors." 2013. arXiv: 1306.6329.

[5] J. Brau et al., "The SiD Digital ECal Based on Monolithic Active Pixel Sensors." Instruments. 6 (4), 51 (2022).

[6] J.S. Marshall and M.A. Thomson, "Pandora Particle Flow Algorithm." Proceedings of Calorimetry for High Energy Frontiers 2013. arXiv: 1308.4537. (2013).

[7] S. Marzani et al., "Looking inside jets: an introduction to jet substructure and boosted-object phenomenology." 2022. arXiv: 1901.10342v4.

[8] A. Steinhebel, "Much Ado About Nothing: Searches for Higgs Boson Decays to Invisible Particles." June 2021. https://scholarsbank.uoregon.edu/xmlui/handle/1794/26657.

[9] J. Yan et al., "Measurement of the Higgs boson mass and $e^+e^- \rightarrow ZH$ cross section using $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ at the ILC." April 2021. arXiv: 1604.07524.

[10] P.A. Zyla et al., "Review of Particle Physics." Prog. of Th. and Exp. Phy. (2020).

[11] E. Coleman et al., "The importance of calorimetry for highly-boosted jet substructure." JINST 13 (2018).

[12] C.T. Potter, "Primer on ILC Physics and SiD Software Tools." Eur. Phys. J. Plus. 135, 525 (2020).

[13] W. Kilian et al., "WHIZARD – Simulating Multi-Particle Processes at LHC and ILC." The Eur. Phys. J. C. 71, 1742 (2011).

[14] J. Alwall et al., "MadGraph 5 : Going Beyond." June 2011. arXiv: 1106.0522.

[15] C. Bierlich et al., "A comprehensive guide to the physics and usage of PYTHIA 8.3." March 2022. arXiv: 2203.11601.

[16] S. Agostinelli et al., "Geant4—a simulation toolkit." Nucl.Instrum.Meth.A 506 (2003).

[17] J.S. Marshal and M.A. Thomoson, “The Pandora Software Development Kit for Pattern Recognition.” The Eur. Phys. J. C. 75, 439 (2015).

[18] Scikit-HEP Project, PyJet. May 17 2023. https://github.com/scikit-hep/pyjet.

[19] M. Cacciari, G.P. Salam and G. Soyez, “FastJet User Manual.” Eur. Phys. J. 72 (2012). arXiv: 1111.6097.

[20] M. Boronat, “Jet reconstruction at high-energy electron–positron colliders.” The Eur. Phys. J. C. 78, 144 (2018).

[21] K. Abe, “Measurement of the branching ratios of the Z0 into heavy quarks.” March 2005. arXiv: 0503005.

[22] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework.” Nucl. Instrum. Meth. A 389 (1997).

[23] A. Hoecker et al., “TMVA - Toolkit for Multivariate Data Analysis.” March 2007. arXiv: 0703039.