SIMULATING OUTPUTS FROM THE HIGH-LUMINOSITY LARGE HADRON COLLIDER FOR IMPROVED ATLAS TRIGGER ALGORITHM TESTING

by

RYAN STUVE

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

November 2023

An Abstract of the Thesis of

Ryan Stuve for the degree of Bachelor of Science

in the Department of Physics to be taken June 2024

Title: Simulating outputs from the High-Luminosity Large Hadron Collider for improved ATLAS trigger algorithm testing

Approved: *Stephanie Majewski, Ph.D.*

Primary Thesis Advisor

The Large Hadron Collider is a proton-proton collider located at CERN near Geneva, Switzerland. It will undergo upgrades starting in 2026 to increase the amount of particle collisions produced by tenfold. This surge in data requires a more effective triggering system within the ATLAS detector located on the collider, so that only interesting collisions are stored for further analysis. This system receives full granularity detector data at 40 MHz, which is then processed by a series of algorithms. Finally, a trigger decision is made to keep or discard each event. One of these algorithms reconstructs topological clusters of energy from individual calorimeter cells. To study the performance of this algorithm, realistic physics simulations are used. The sample analyzed in this paper is simulated productions of two Higgs bosons, a rare event that could provide more insight into the properties of the Higgs boson. By reformatting the data of the sample into detector readout, the algorithm can be improved to better recognize di-Higgs events against a background of less interesting particle interactions. This thesis will describe the reformatting process in detail, as well as performance tests used to validate the simulated data and study clustering algorithms.

Contents

1 Background and Motivation

1.1 Background: What do we know so far [2-3 pages]

In 430 BCE, Greek philosopher Democritus claimed the universe was composed of tiny building blocks, which he termed “atomos."[9] Since then, scientists have employed the properties and interactions of atoms to explain phenomena at microscopic scales. In 1917, Ernest Rutherford discovered the proton, a particle smaller than the size of an atom. The proton was soon complimented by the neutron, motivating a new field of subatomic physics. As more subatomic particles were discovered, patterns of behavior began to emerge.[18] One of these patterns included a symmetry in mass between the proton and neutron, despite their charge difference. To address this phenomena, a new set of particles known as ’quarks’ was proposed, which combined to form protons, neutrons, and other heavier particles. Decades of research culminated in the 1970s with the creation of the Standard Model, physicists’ best template for describing the elementary particles that make up the universe.[21] In the next few decades, the top quark (1995)[1] and Higgs boson (2012)[4], particles predicted by the model, were experimentally observed.

The Standard Model splits elementary particles into two categories, bosons and fermions. (See Fig. 1) The bosons are carriers of three of the four fundamental forces. Photons carry the electromagnetic force, gluons carry the strong force, and W and Z bosons carry the weak force. At the time of writing, no corresponding particle for gravity has been observed. Fermions are the building blocks of matter, divided into quarks and leptons. The up and down quark are the lightest, most stable quarks, and combinations of the two create the proton and neutron. The up quark has two heavier sibling particles, the charm and top quark, that have identical properties but much higher masses. (2.2, 1280, and 173100 MeV/c^2 respectively) These are called the generations, and each type of fermion has three. For example, the down quark is the first of three particle generations, accompanied by the strange and bottom quarks. (4.7, 96, and 4180 MeV/c^2)

The electron also belongs to a three-generation lepton family, including the muon and tau. Just like the quarks, higher generations in this family have significantly higher masses. The final family of leptons is the neutrinos, which are extremely light and rarely interact with other particles. Because of their elusiveness, the behavior of neutrinos is difficult to measure, and serves as a possible stepping point into a theory beyond the Standard Model.

Interactions, or couplings, between the families of fermions are mediated by bosons, and can be predicted by properties described in the Standard Model, such as mass and charge. For example, neutrinos cannot interact with the electromagnetic force because they have no charge, and fermions can change families by

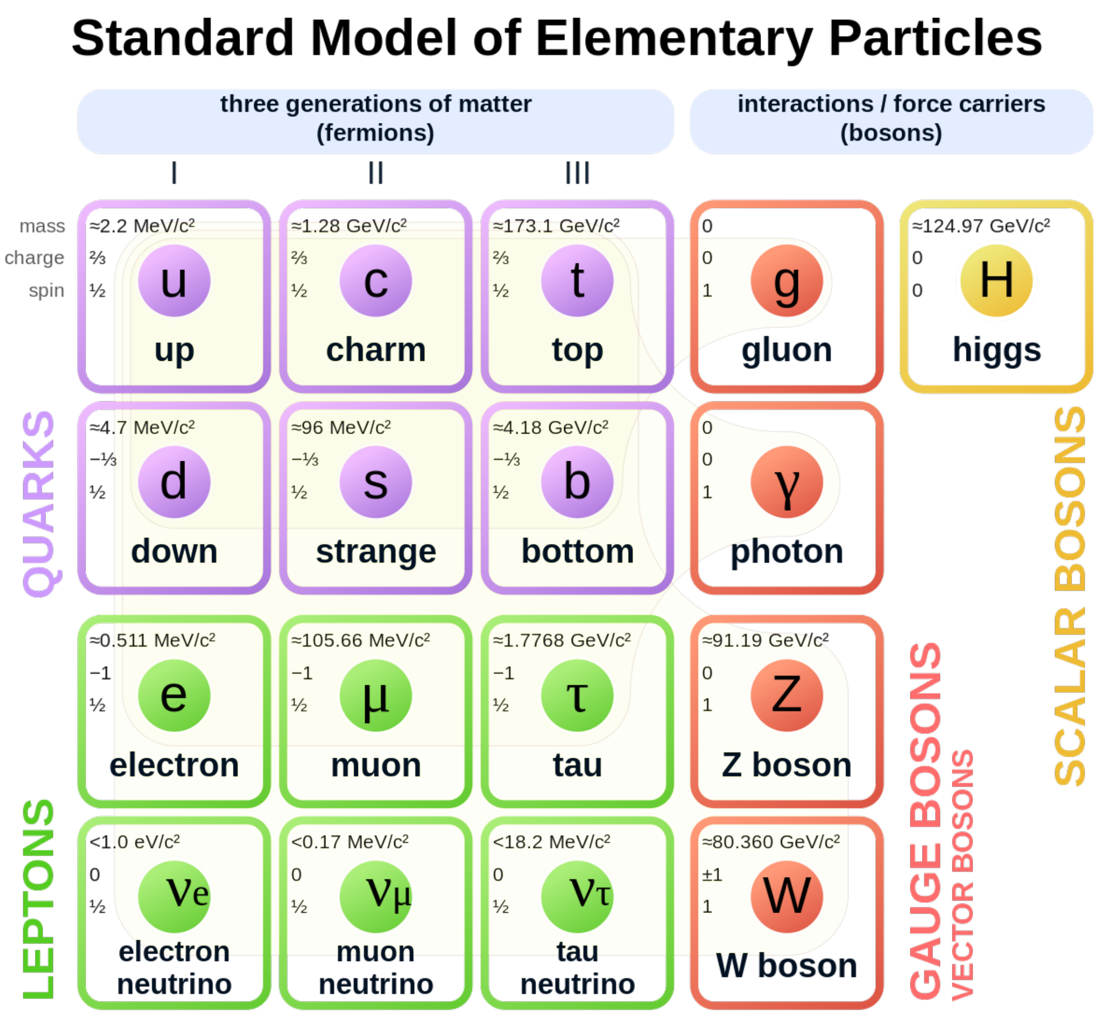

Figure 1: Particle content of Standard Model. Fermions are on the left in their respective families, bosons are on the right. Values are subject to adjustment. Source: Wikipedia: 'Standard Model of Elementary Particles' by MissMJ

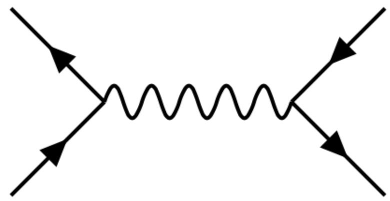

Figure 2: Example of a Feynman diagram. Solid lines represent fermions, while the wavy line represents a photon or gluon. If the time axis goes from left to right, this represents a fermion annihilating with its anti-particle to produce a photon, which then decays into a particle anti-particle pair.

decays through the weak force, emitting a charged W boson.

In 1918, Emmy Noether published one of the most important papers to modern physics. This paper detailed what is now known as Noether's theorem, stating that every symmetry of the universe corresponds to a conservation law.[25] For example, time symmetry claims that two identical systems will behave the same way regardless of the time frame in which they exist. This symmetry implies the conservation of energy between time frames. Other symmetries include rotational symmetry, which gives rise to conservation of angular momentum, and charge conjugation symmetry, which corresponds to conservation of charge. Only in extreme situations like black holes do these symmetries break down.

All particle interactions are governed by fundamental symmetries, as visualized through a notation known as Feynman diagrams.[16] (See Fig. 2) By imposing strict limits on how each particle can couple, this notation indicates which reactions are most likely to take place and how they occur. One unique feature of particle physics is the Heisenberg uncertainty principle, which places a lower limit on measurement accuracy. It states that $\Delta E \Delta t > \hbar/2$, meaning a small time uncertainty allows for a greater uncertainty in energy.[12] Since energy and mass are directly related, this allows particles to temporarily violate conservation of mass, with greater mass violations requiring smaller time scales. In other words, particles can vary from their rest mass as described by the Standard Model, but must quickly decay into stable particles. These "virtual particles" cannot be directly detected, but their effects lead to various observable phenomena such as quantum fluctuations. It also allows lighter particles like the up quark to emit W bosons, despite their rest masses differing by a factor of over four million.

Another way to understand particles is as excitations of fields. There exists fields for each fermion and boson which are allowed to interact with each other.[27] Interactions with a boson field manifest themselves as effects of the corresponding fundamental force. Physicists can predict the probabilities of particle couplings, referred to as cross sections, and confirm them through experiments. Particle accelerators like the Large

Hadron Collider are especially useful for creating particles with smaller cross sections. They do this by accelerating charged particles through electromagnetic fields, then colliding them into other stationary or moving particles.[11] (See section 2.1) These collisions generate huge amounts of energy that then produce massive particles like the Higgs boson and top quark. Since heavy particles are unstable, they quickly break down into multiple, more stable particles. The ratio at which these particles decay into specific products is called the "branching ratio," and provides insight into the properties of the particles.[10] A goal of high-energy experiments is to collect data about the decay products of less easily-produced particles like the Higgs boson.

The data that has been collected on the Higgs boson suggests that all particles obtain their relative mass based on how strongly they interact with the Higgs field. This observation matches predictions of the Higgs mechanism, a theoretical process that explains the unique properties of the electromagnetic and weak forces. According to the Higgs mechanism, above a certain energy threshold (about 160 GeV) the two forces behave identically. As the energy drops below this value, the particles that make up these forces interact differently with the Higgs field.[2] The photon does not interact at all and therefore remains massless, while the W and Z bosons strongly interact and assume GeV-scale masses. With each new field must exist a new particle, so the observation of the Higgs boson provides further confirmation of this mechanism. More extensive probing of the boson's properties could give insight into the Higgs mechanism, or answer unexplained features of the Standard Model such as the large variance between elementary particle masses.

1.2 What is missing in our physics knowledge? [2-3 pages]

The Standard Model is capable of making extremely accurate predictions about particle properties. For example, it predicts the magnetic moment of an electron to within a billionth of a percent of the measured value.[15] Despite its impressive accuracy, the model still fails to explain various experimental phenomena. One example mentioned earlier is unexpected neutrino behavior. The original Standard Model predicted neutrinos to be massless, yet various experiments have suggested they do have some small but measurable mass.[22] Another issue unexplained by the current Standard Model is baryon asymmetry, why there is more matter than antimatter in the universe.[24] Dark matter, an unobserved 85 percent of the mass in the universe, remains unaccounted for by the Standard Model.[3] Similarly, dark energy, which has been proposed as an explanation for the accelerating expansion of the universe, is left unexplained.[23] A final important missing piece is an explanation for gravity. While Einstein's general theory of relativity successfully predicts the motion of large bodies like galaxies, it breaks down at the smallest scales of matter, preventing the existence of a Grand Unified Theory for the four fundamental forces.[13]

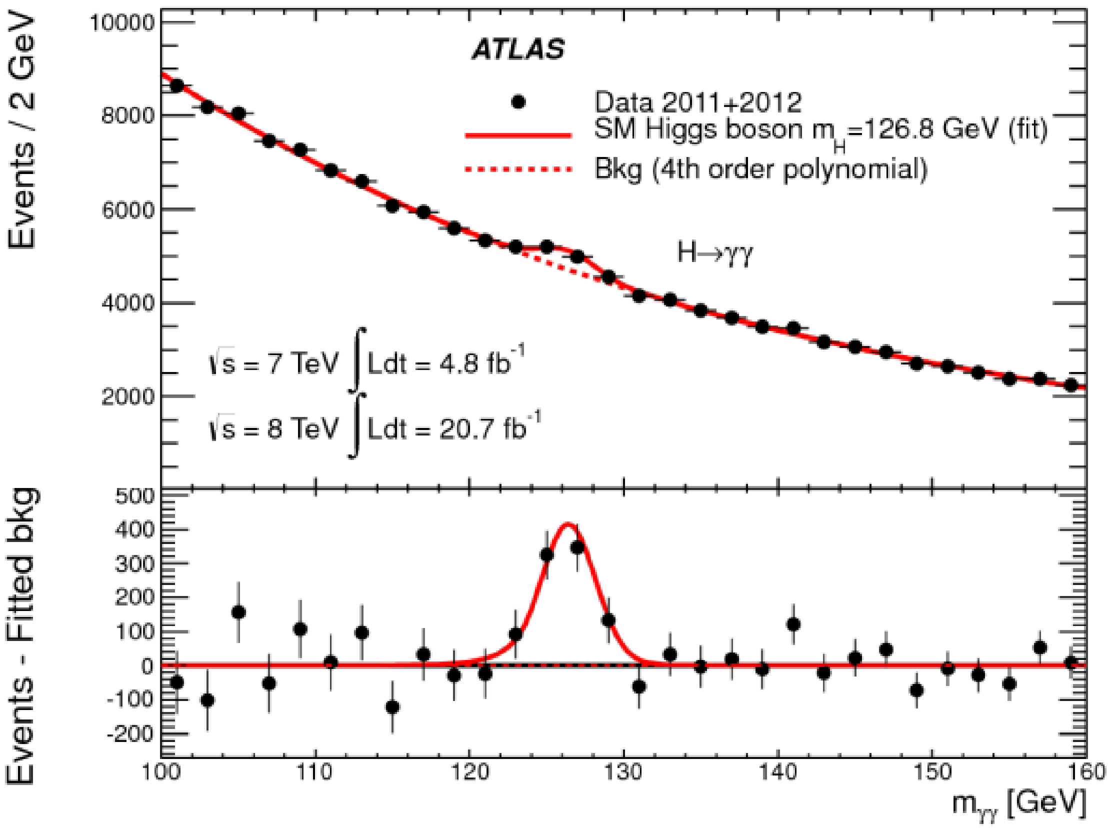

Figure 3: A plot used to validate the discovery of the Higgs. By combining information from two photons in an event, the mass of their parent particle can be estimated. A peak occurred in the data around 125 GeV, implying the parent particle had a tendency to be that mass. When the peak exceeded a 5 sigma difference from background, the Higgs boson's existence could be confirmed. Source: CERN

An effective way of discovering new physics is through testing predictions from the Standard Model. If measurements disagree with the model, it suggests areas in the theory where modifications should be made. The Higgs boson was discovered in 2012 using such an analysis. Choosing a specific set of decay products (two photons), physicists estimated the distribution of their combined mass over many events. This estimate was based off the Standard Model and predicted collider behavior, and formed a background hypothesis. As data was collected and compared to background, a significant peak began to emerge at 125 GeV, the Higgs boson rest mass. (See Fig. 4)

A measurement physicists are currently pursuing is the rate of Higgs boson self-coupling. The Higgs potential, a fundamental part of the Higgs mechanism, requires the Higgs boson to interact with itself at a specific cross section, about 36.69 fb at 14 TeV.[19] Using high-energy particle accelerators, collisions that produce two Higgs bosons can be measured and analyzed. (See Fig. 3) If the frequency of these di-Higgs

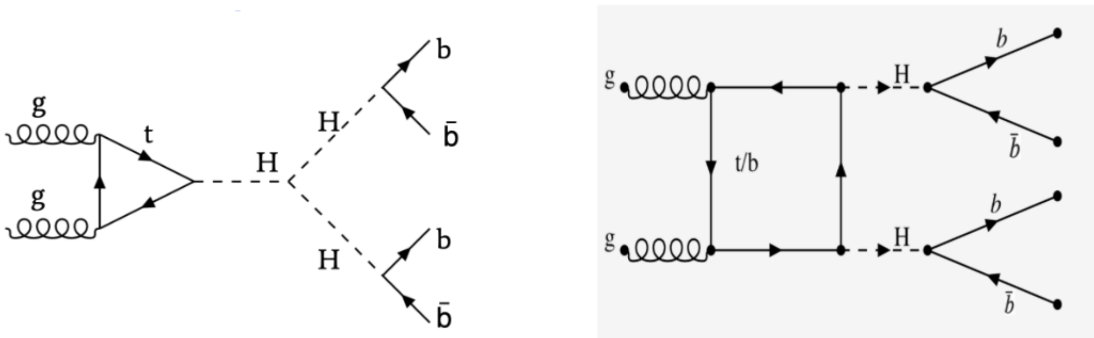

Figure 4: Possible Feynman diagrams for the production of two Higgs bosons within CERN's Large Hadron Collider. Read from left to right, two gluons collide within a detector, producing a top or bottom quark. This high energy quark quickly decays to two Higgs bosons. Before reaching the detector, the bosons also both decay, in this example to a bottom quark and anti-bottom quark. The decay is often abbreviated to HH4b, meaning 2 Higgs decaying to four b quarks.

events and their decay products do not match the prediction of the Standard Model, it could mean more particles exist, or our understanding of the known particles needs to be modified.

The Higgs boson has a large enough cross section that its production events are capable of standing out from background. Di-Higgs events have a significantly smaller cross section, meaning they do not occur at a high enough rate to generate a noticeable peak in the current physics data. Only by increasing the amount of events and removing unimportant particle interactions can any statistically significant trends be observed.

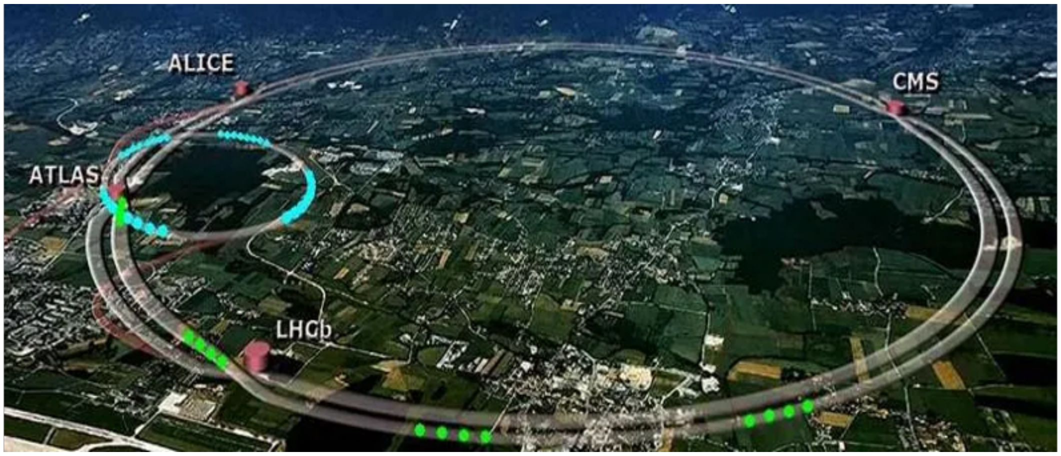

Figure 5: Above-ground view of LHC, with detectors. Source: CERN

2 Methods

2.1 HL-LHC

The Large Hadron Collider (LHC) is a high-energy particle collider located near Geneva, Switzerland. Since construction in 2008, the collider has completed two successful data-collecting runs.

The LHC primarily operates as a proton-proton collider. Bunches of protons are accelerated in both directions through a twenty-seven kilometer circular ring to 99.9999991 percent the speed of light.[14] The bunches are then brought together and collide at four specific points along the ring. (See Fig. 5) For Run 3 specifically, these collisions take place at a center-of-mass energy of $\sqrt{s} = 13.6$ TeV, creating about fifty interactions every twenty-five nanoseconds. These interactions emit high-energy particles into surrounding detectors, which then extract information from them. The four collision detectors are ATLAS, CMS, ALICE, and LHCB.

At the time of writing, the LHC is in its third run and expects to collect roughly 200 fb^{-1} of data. After completing this run, the collider will enter a shutdown and upgrade period from 2026 until 2029.

The third long shutdown (LS3) will enhance the production and collection capabilities of the collider by tenfold. By improving the luminosity (a measurement of the rate of proton collisions in the LHC), over 3000 fb^{-1} of data can be provided for collection in the detectors over a period of twelve years.[20] For this reason, the upgraded collider will be renamed the High-Luminosity Large Hadron Collider (HL-LHC), and will collide protons at a center-of-mass energy of $\sqrt{s} = 14$ TeV.[17]

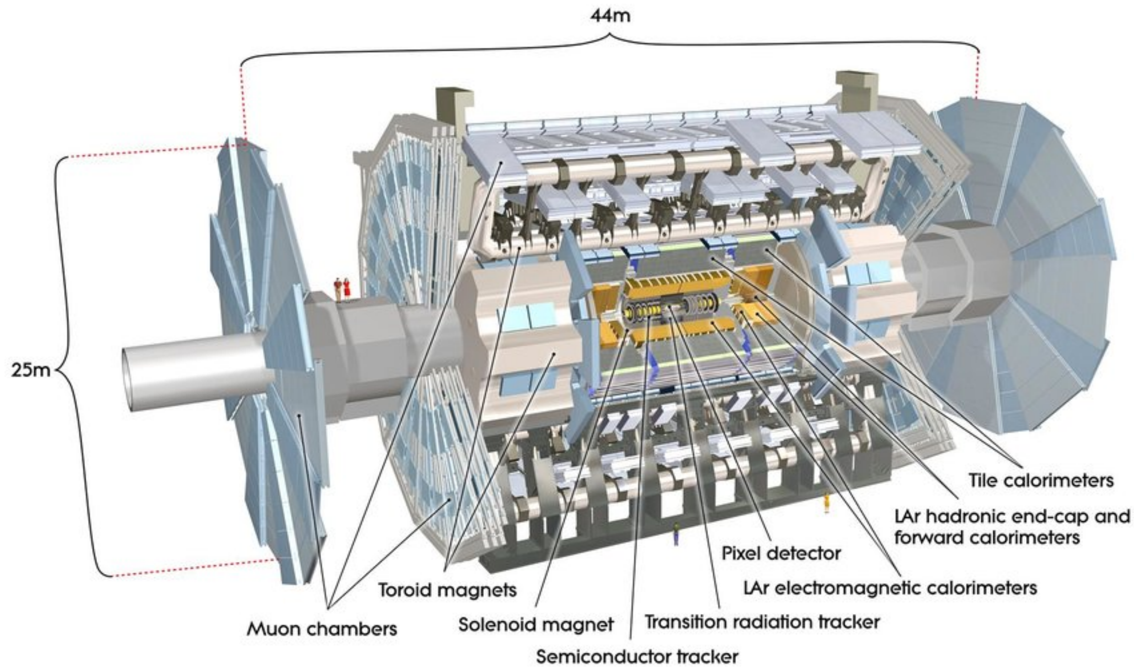

Figure 6: Depiction of ATLAS, with detector systems labeled. Source: CERN

2.2 ATLAS Detector

The ATLAS detector is a multi-system instrument used in the detection of particles ejected from the first collision point of the LHC. It serves as an experimental device for measurements of the Higgs boson and top quark, as well as searches for dark matter. Over a billion collisions take place in the ATLAS detector every second, creating a substantial flow of data that must then be processed and stored. Limitations in computational speed require that only one in every million collisions can be saved for further study. ATLAS must analyze and determine whether each bunch crossing, or event, contains potentially interesting data. It does this by interpreting signals coming from six different detection systems, including the pixel detector, calorimeters, and muon chambers.[7] Large magnets on the detector help to orient charged particle trajectories in order to ensure accurate momentum measurements. (See Fig. 6)

The calorimeter system exemplifies one method of translating real particles to computer data. Particles from collisions shower through this system as they are stopped, completely depositing their energy into 184,000 individual cells.[26] Energy and location data can be reconstructed to provide details about the original collision and particles produced.[5]

Simulated data is a valuable tool for ATLAS. Using current understandings of particle physics, such as the standard model, Monte Carlo simulations can create events and output realistic detector data. This data can be useful as an expectation value in model validation studies, or as sample data in detector upgrade tests.[8]

2.3 Trigger & Data Acquisition System

The high rate of collisions at the LHC requires ATLAS to efficiently collect data from its detector systems. This is especially true for the HL-LHC runs, which expect a bunch crossing rate of 40 MHz.[6] Using current software, this rate would produce a volume of data on the exabyte scale (1,000,000 terabytes). ATLAS does not have the capacity, nor the need, to collect such a large dataset. Instead, it employs a trigger system to filter out more common events and reduce the amount that get saved to permanent storage.

To do this, the Trigger and Data Acquisition (TDAQ) system imposes a list of criteria that incoming detector data must pass. The criteria is motivated by current physics searches, such as Di-Higgs production and supersymmetry (SUSY). These processes are predicted to produce specific signatures within the detector that the trigger system can identify. Among these signatures is an isolated lepton above a certain transverse momentum, which can signal Higgs or SUSY particle production. Di-photon signatures also correspond to Higgs and SUSY particles. Di-jet signatures are especially useful, as they signal Higgs self-coupling events, one of the primary motivations of the HL-LHC. The same signatures can also be produced through other processes, meaning any event containing such a signature is only a candidate for the process being studied.

The process of collecting data from the detector systems is extremely complex. Each individual cell must have a wire that transmits data to the trigger. Because the rate of collisions is so high, the trigger has only a few microseconds after receiving the final signal from an event before the closest signal from the next event arrives. Within this time, the trigger must run algorithms to reconstruct any significant features, compare them to relevant signatures, and return a decision to keep or discard the event. When the trigger accepts a certain event, it transmits the decision to the detector readout system, which then saves the data from each individual cell for that event into permanent storage.

The output rate is 10 kHz, four thousand times smaller than the collision rate. While reducing the amount of data stored, the trigger system also promotes collecting an abundance of interesting events, while suppressing more common and well studied events.

As HL-LHC upgrades begin, the TDAQ system will be redesigned to handle the expected changes in luminosity. These Phase-II upgrades will allow the system to implement more efficient and detailed algorithms for identifying features in the detector data. (See Fig. 7)

The new TDAQ system will be able to make selections based on not only momentum and multiplicity,

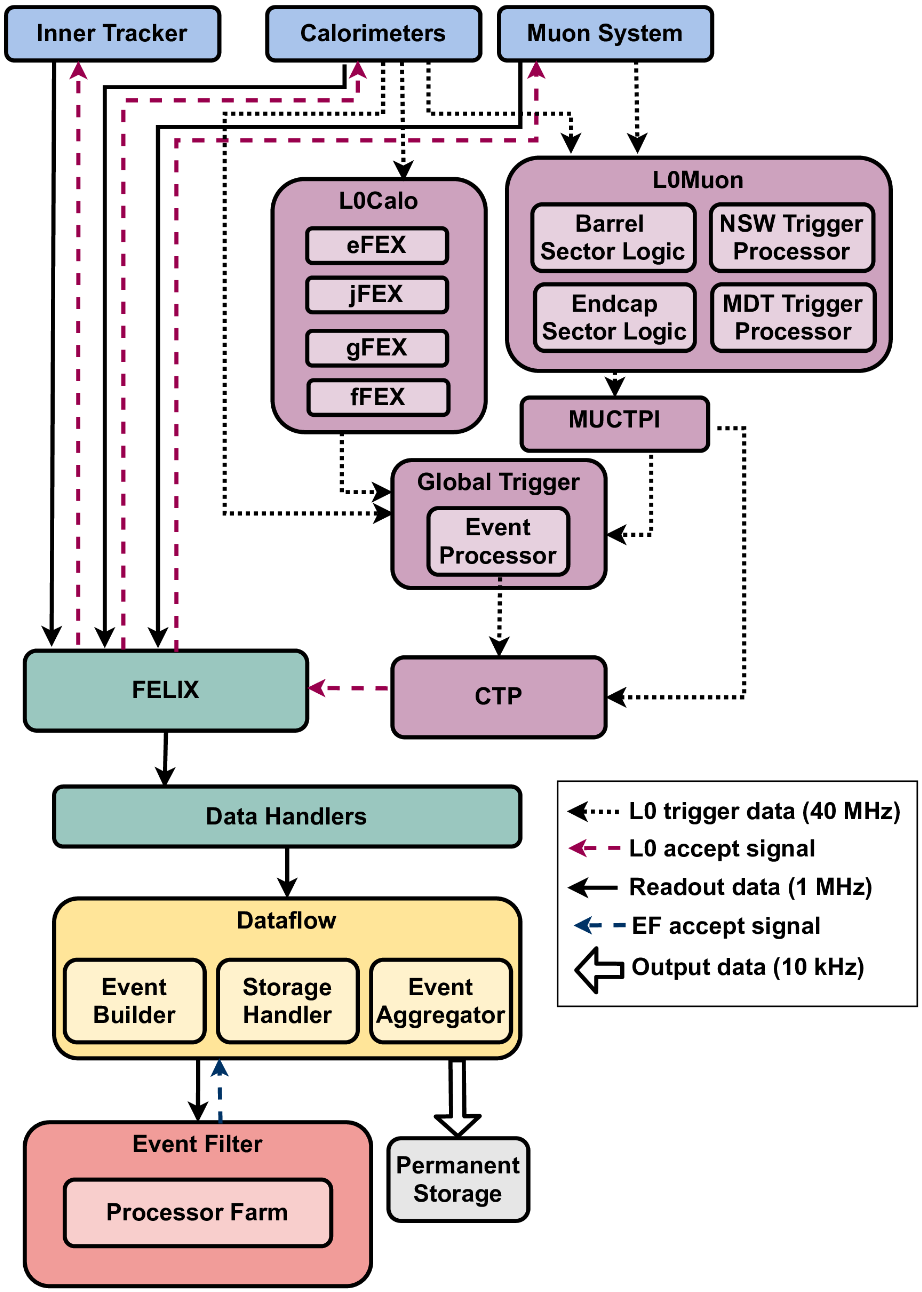

Figure 7: Diagram of upgraded ATLAS Trigger and Data Acquisition system, with relevant rates of transmission labeled. Source: ATLAS TDAQ Phase-II Upgrade Technical Design Report

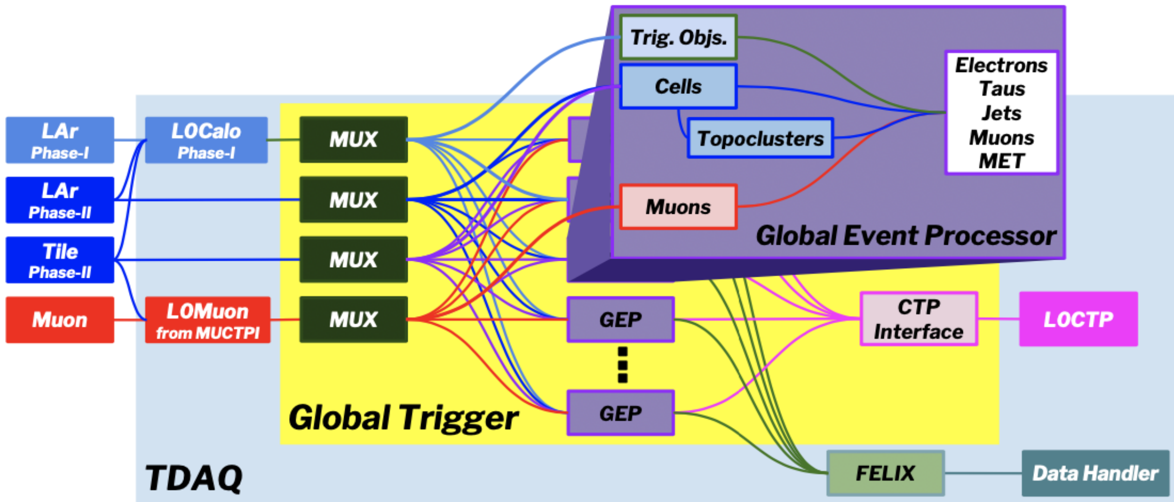

Figure 8: Global Trigger system. Lines represent data flow. Source: ATLAS TDAQ Phase-II Upgrade Technical Design Report [6]

but also the location (topology) within the detectors.

2.4 Global Trigger

One of the new features of the Phase-II TDAQ system is the Global Trigger. This component will use custom algorithms to process information from both the calorimeter and muon detectors. Before entering the Global Trigger, data goes to the L0Calo and L0Muon subsystems, which create Trigger Objects (TOBs). TOBs from L0Calo are created by applying feature extractors to coarse granularity calorimeter data. They include jets, leptons, photons, and missing transverse energy. TOBs from L0Muon only include candidates from the muon detectors.

Multiplexers within the Global Trigger route the TOBs and full-granularity detector data into a singular Global Event Processor (GEP). Each GEP is housed on a Field Programmable Gate Array (FPGA), which allows for flexibility and optimization. The GEP uses calorimeter data to subtract pile-up, create topological clusters, and identify various topological signatures. This information is combined with the TOBs from L0Calo and L0Muon to refine the object selections. Once the TOBs are refined, they are passed through algorithms that search for interesting signatures within the data. Finally, the TOBs are transferred to the Central Trigger Processor, which sends the signal to keep or discard an event. (See Fig. 8)

The entire process of obtaining and analyzing data must occur within the Global Trigger for each individual bunch crossing, an endeavor that had not been done in previous runs of the collider. At a rate of 40 MHz, this allows .000001 seconds for the trigger to handle data from one event before receiving the next. To overcome this time limitation, the system uses time multiplexing to process several events in parallel. After the data from one event is sent to a GEP, the multiplexers shift their output to the next. Only after all 48 GEPs have received data from an event will the multiplexers restart their indexing. This method grants more time for each GEP to properly process events, so that the entire Global Trigger can handle an incoming 50 TB of data per second.

2.5 Topoclustering Algorithm for the Global Trigger

The upgraded TDAQ system will use full granularity data from the calorimeter to precisely locate energy deposits for an event. Within the Global Trigger, a clustering algorithm will create topological clusters (topoclusters), which are used to reconstruct particles and measure missing energy. A single cluster is not expected to represent the complete detector response to a particle. Instead, the location and energy of multiple clusters are recombined using an anti-kt algorithm to provide signatures for the trigger system.

Clusters are created by adding neighboring cells that contain energies above a certain threshold. This threshold uses a combination of electronic noise and pile-up noise that occurs as remnant energies from previous collisions are read into the current event. A cell reading more than two times this energy value is called a “2-sigma cell", while one with more than four times the value is called a “4-sigma cell". The 420 and 422 algorithms take advantage of these distinctions to build clusters. Each starts with a 4-sigma cell (seed cell), then adds any three-dimensional direct neighbors with energies over 2 sigma.[1] The neighbors of these newly added cells are then checked and added if they pass the 2-sigma threshold. This process continues until there are no more 2-sigma neighbors to append, at which case the 422 algorithm will be complete. The 420 algorithm differs in that it ends by adding all neighboring cells of the final 422 cluster with energies above 0 sigma. (See Fig. 9)

These two cluster algorithms are offline algorithms, meaning they are performed on data that has already been stored from the detector to memory. Therefore, there is no limit to the time complexity of the algorithms. Furthermore, limited bandwidth from the calorimeters prevents energy readings below 2 sigma from being fed to the Global Trigger, making a 420 algorithm impossible to implement in the detector. The upgraded TDAQ system will require an online algorithm to perform similar functionality at the rate of the Global Trigger. One such algorithm, the Unique Seed Cell Association (USCA) algorithm, works by limiting the

[1]The ATLAS detector uses a common coordinate system with the origin placed at the interaction point, and the z-axis running along the beam line. The azimuthal angle ϕ measures the angle around the beam, while the polar angle is reported in terms of pseudorapidity $\eta = -\ln \tan(\theta/2)$

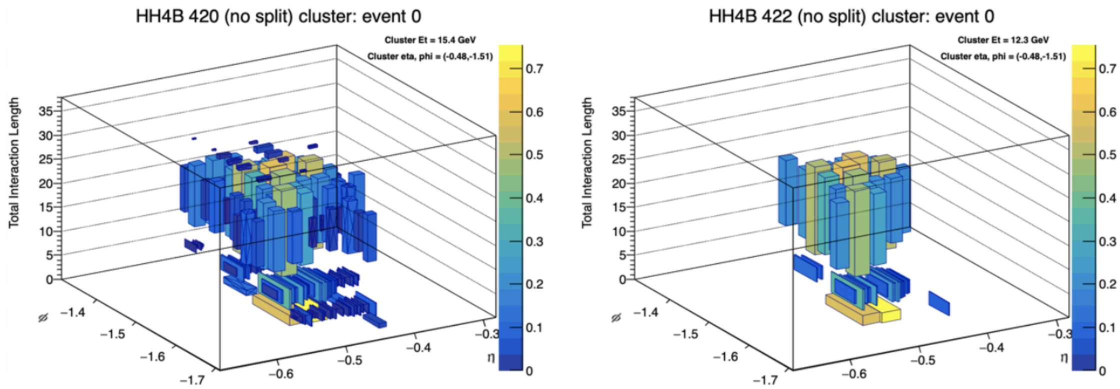

Figure 9: Visuals of topologically identical clusters formed by the 420 and 422 algorithms. Cell volumes are scaled to their corresponding transverse energy. The 420 cluster is surrounded by low energy cells that are appended after the 422 algorithm is complete. (Code adapted from Sylvia Mason)

number of iterations (shells) performed by the 422 algorithm. (See Fig. 10)

Monte-Carlo simulated detector data is being used to test the performance of this algorithm within the HL-LHC, as well as validating implementations on the FPGAs. Comparing the USCA algorithm to offline algorithms determines what variations of the algorithm (1-GeV cuts, number of shells) make it most effective for triggering on interesting events. (See Fig. 11)

To predict the expected amount of relevant data, analysis was performed on simulated data for HH4B and JZ0W processes. HH4B is the most important process for the clustering algorithm to identify, and JZ0W provides an example of background, where no significant processes occur. More specifically, JZ0W simulates the effects of strong force interactions from jets within the detector. The average and maximum percentage of cells above 2 and 4 sigma for each sample set expectation values for the running HL-LHC. (See Fig. 12,13)

3 Firmware Validation

At the time of writing, firmware is being developed to implement online topoclustering on FPGAs. Validation studies must be performed with each step to ensure the algorithm will behave as expected. These studies use simulated HL-LHC data as input to resemble the conditions of the upgraded collider. (Section 2.2) To do this, the data must be formatted as though it were being received directly from the detector systems.

One of the projects included in this thesis is an algorithm that takes data from higher-level files, known as root files, and reformatting them into lower-level .coe files resembling detector readout. Root files contain lists of information (ntuples) such as cell locations and energy readings for a given number of events. The root_to_coe algorithm works by parsing ntuples one event at a time, using the data to create a map of

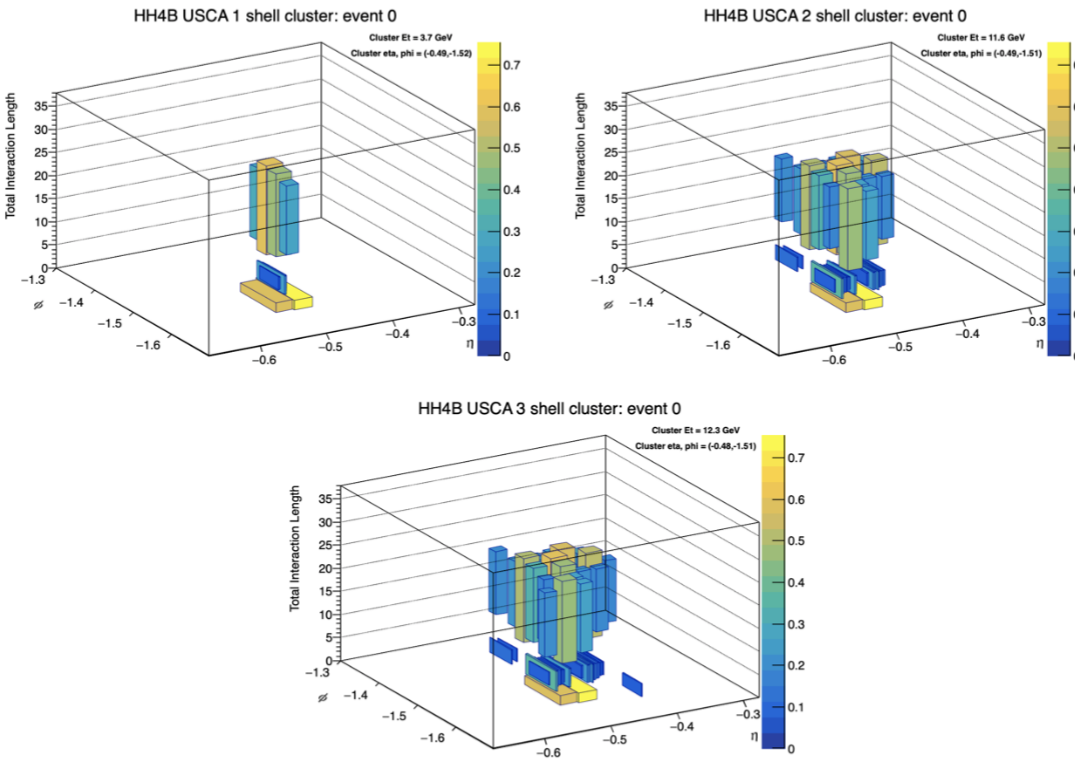

Figure 10: USCA cluster after different numbers of iterations. Each iteration appends 2-sigma cells that neighbor cells already belonging to the cluster. Cluster is topologically identical to Figure 5.

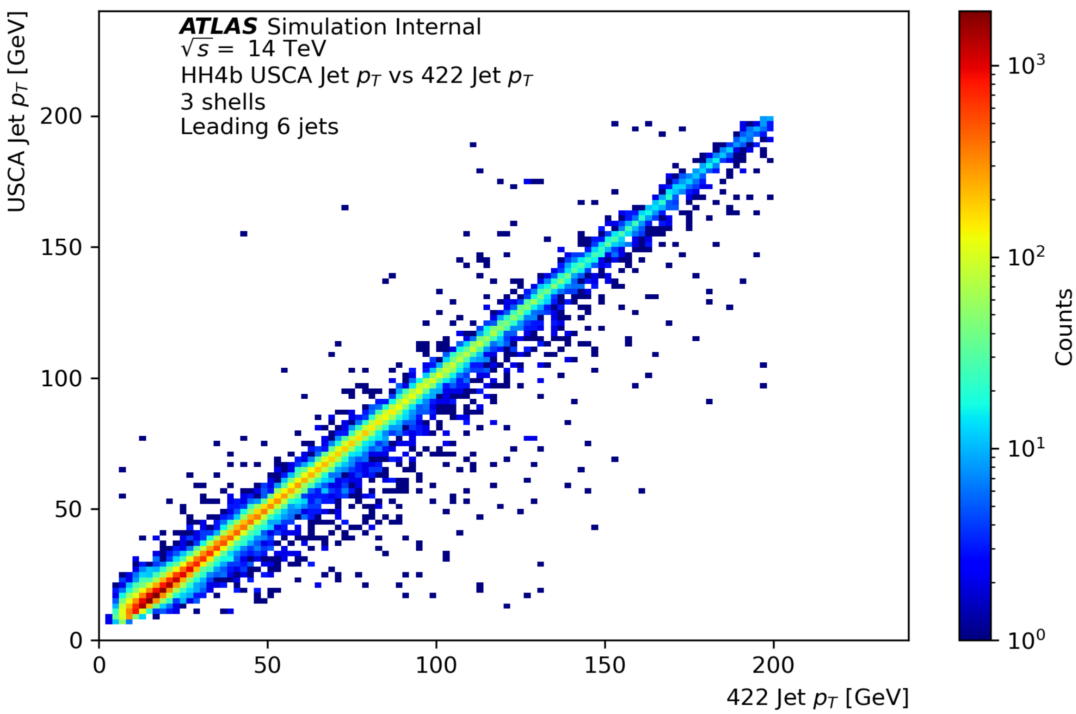

Figure 11: Comparison of the performance of USCA algorithm compared to 422 offline algorithm. For this plot, the 6 highest p_{T} jets generated by each algorithm were taken for a simulated HH4b sample, and their reconstructed p_{T} were compared. The resulting line demonstrates strong agreement between the two algorithms. Source: Tim Mathew, University of Oregon

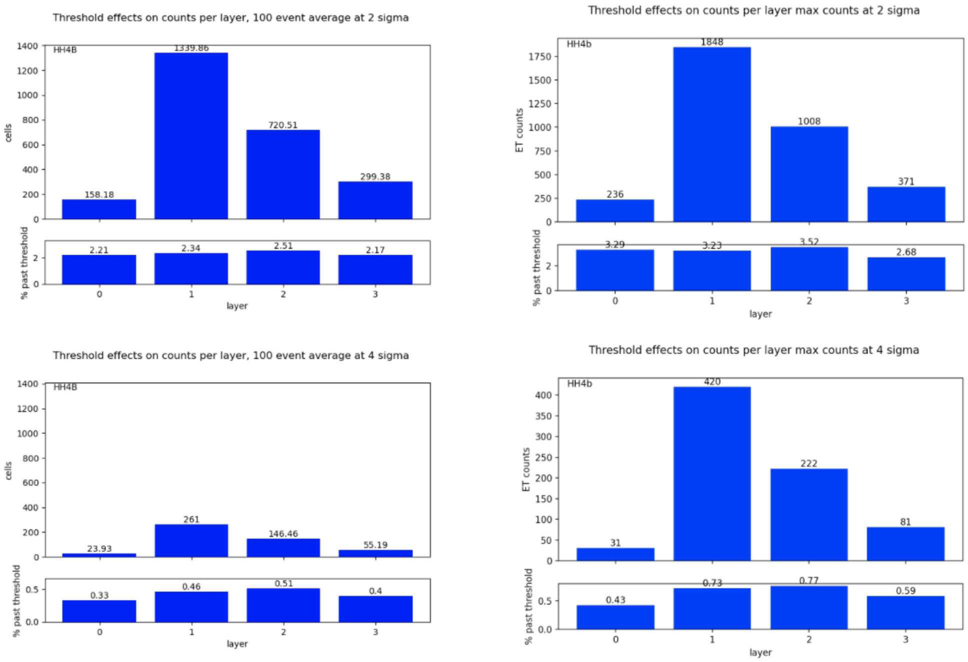

Figure 12: Threshold effects on first four layers of Electromagnetic Barrel for simulated HH4B sample, showing percentage of cells past 2 sigma on top and 4 sigma on bottom. On the left is values over a 100 event average, on the right is the maximum values for each layer within the 100 events. HH4B events on average have roughly 2.5% 2-sigma cells and 0.4% 4-sigma cells.

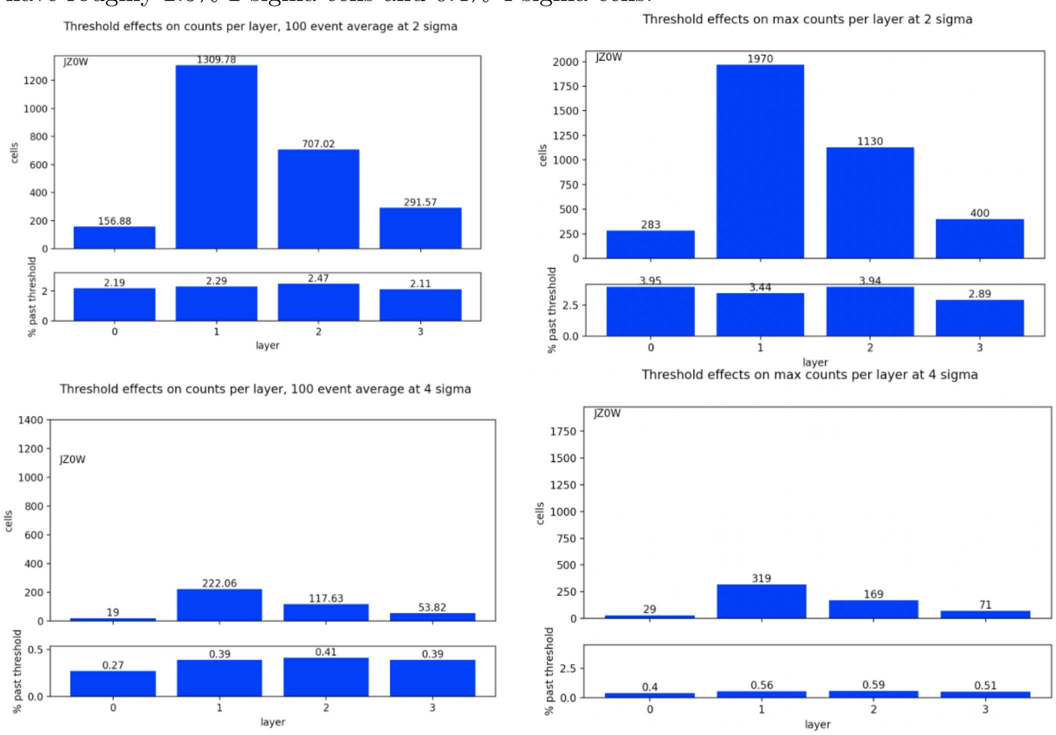

Figure 13: Threshold effects for simulated JZ0W sample. JZ0W events on average have roughly 2.3% 2-sigma cells and 0.35% 4-sigma cells.

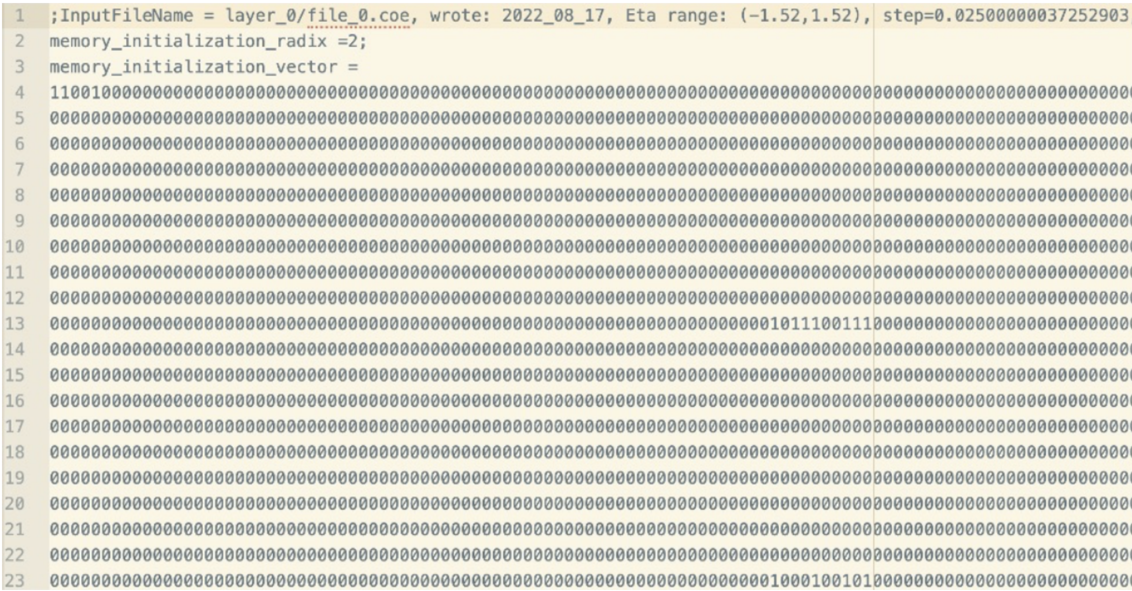

Figure 14: A .coe file, including the header. Each 10-bit number represents one cell's energy reading for that event. Rows correspond to eta slices, and columns to phi slices. Events are stacked vertically in the file.

the detector. This map has two dimensions, eta and phi, and resembles an unrolled layer of the detector. The rows are "slices" of a common eta, starting at the most negative value. In the row are cells sorted by phi value from $-\pi$ to π. At each cell's location, a binary value is written describing the recorded transverse energy. Once all energy values have been recorded for an event, the map is saved to a .coe file. (See Fig. 14)

The next event is written below, and so on until the specified number of events have been reformatted. The root_to_coe algorithm outputs one .coe file for each layer, along with information about the dataset stored in text files.

There are various functionalities within the algorithm that can be modified. One such function is the ability to introduce a threshold to the energy values. This threshold is a multiple of sigma (2 or 4 in the case of topoclustering), and any energy below it is set to zero. Another useful function is the ability to suppress bad cell readout. Provided a database of cell quality information, the algorithm can identify which cells are malfunctioning and set their energy values to zero for all events.

Along with the root_to_coe algorithm, a program was written to help visualize the output files. This program takes a .coe file and a specified event and plots it onto a two-dimensional heat map, with the colors representing the transverse energy values. (See Fig. 15)

Visualizing the .coe files is useful not only for understanding what is happening within the detector, but also identifying and analyzing potential clusters within a single layer.

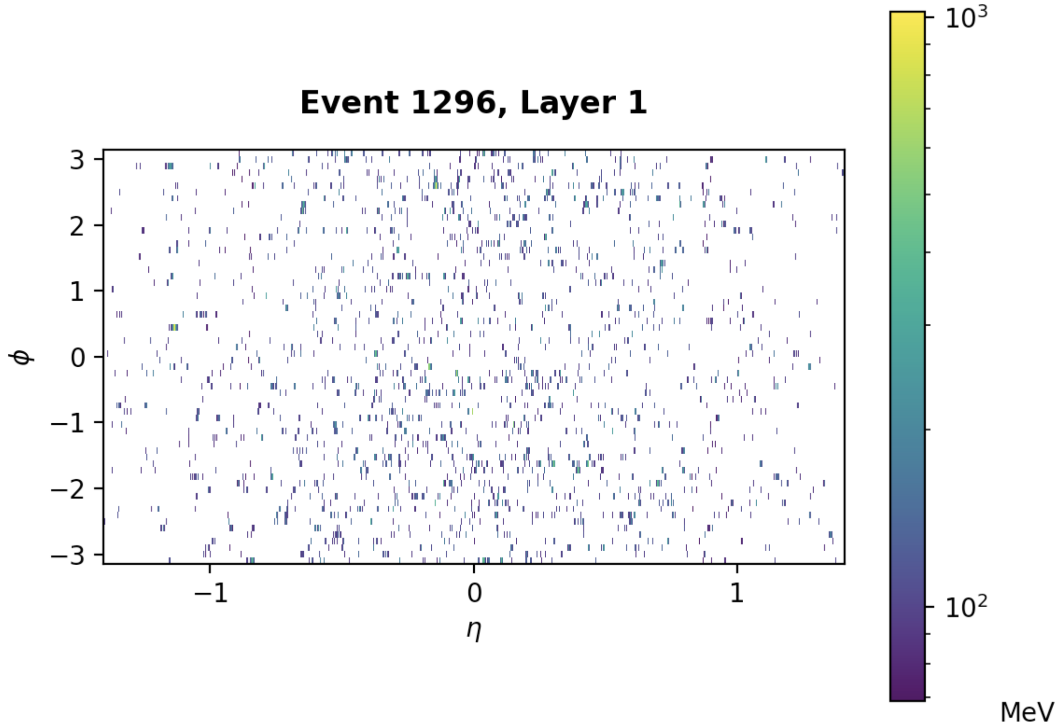

Figure 15: A visual interpretation of the .coe files produced by the root_to_coe algorithm. Colors represent the magnitude of recorded energy in the cell.

4 Di-Higgs Performance Studies

Another project included in this thesis is the reconstruction of the Higgs mass from a simulated Di-Higgs sample. Each boson can decay into various combinations of particles, but most often both decay into a bottom and anti-bottom quark pair. This specific decay occurs for about a third of all di-Higgs productions. Bottom quarks have a unique way of showering into detectors, creating identifiable jets. In this sample, all events contain two Higgs decaying into 4 bottom quarks, which is commonly referred to as an HH4b event. If the data behaves as expected, it should be possible to identify the two jets for each Higgs and combine them in a way that produces properties consistent with the parent boson.

After clusters are formed in the Global Trigger, they are run through an anti-kt algorithm to form jets. The Di-Higgs sample contains jet reconstructions using all three clustering algorithms: 420, 422, and USCA. Because both the clustering and anti-kt algorithm are generalized, they are prone to error. This error can lead to jet reconstructions at incorrect energies or locations. One benefit of using simulated data is the existence of “truth data". Truth data provides the most accurate reconstruction of the simulated event, which can be used as a comparison in validation studies. In this particular sample, truth jets provide an ideal set of data to apply the reconstruction algorithm on.

The first step to reconstructing the Higgs is to identify the six jets with the highest transverse momentum (p_{T}) in an event. Since the b quarks should travel in roughly the same direction, their jets should be close in eta-phi space. (See Fig. 16) Thus, the two closest jets in an event are likely to come from one parent Higgs. By only including the six highest p_{T} jets, the probability of selecting two unrelated low-energy background jets is greatly reduced. After these two jets are selected, the next two closest of the remaining four become the second di-jet choice. The USCA algorithm is limited to a range of $|\eta| < 3.2$, so if the four selected 420 or 422 jets are beyond this range, the event is discarded for consistency.

Once identified, the jets can be converted to four-vectors, which contain data about mass and momentum. The four-vectors for each pair are added component-wise to produce a vector for the parent Higgs. This final vector contains information about the Higgs' transverse momentum and mass. Comparing the p_{T} of the two final vectors and selecting the higher one makes it even more likely that the final vector comes from a Higgs boson. The parent mass is then extracted from this vector. If the reconstruction algorithm was able to correctly select a jet pair from an individual Higgs boson, a histogram of mass values should return a distribution with a noticeable peak. This peak represents the reconstructed Higgs mass.

Since truth jets contain exact data, as opposed to the 420, 422, and USCA jets, applying the reconstruction algorithm to truth data should produce the most defined peak. The results from the simulated sample are

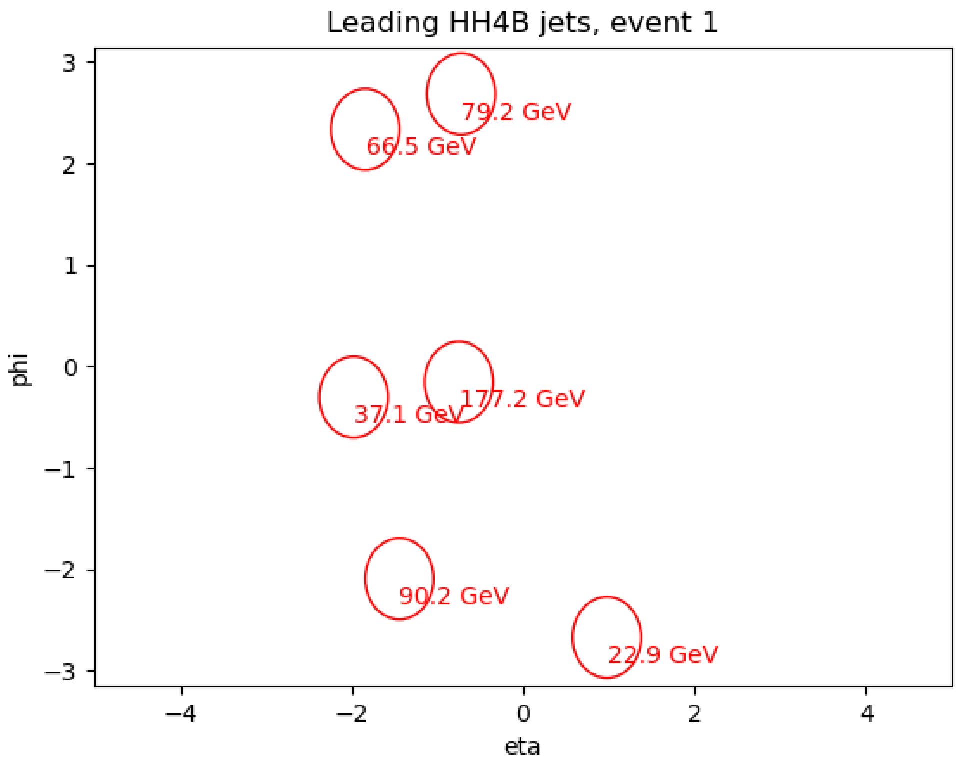

Figure 16: Six highest p_{T} 420 jets from a simulated Di-Higgs event. The two closest pairs of jets are reconstructed as bottom quark pairs that emerged from a Higgs boson.

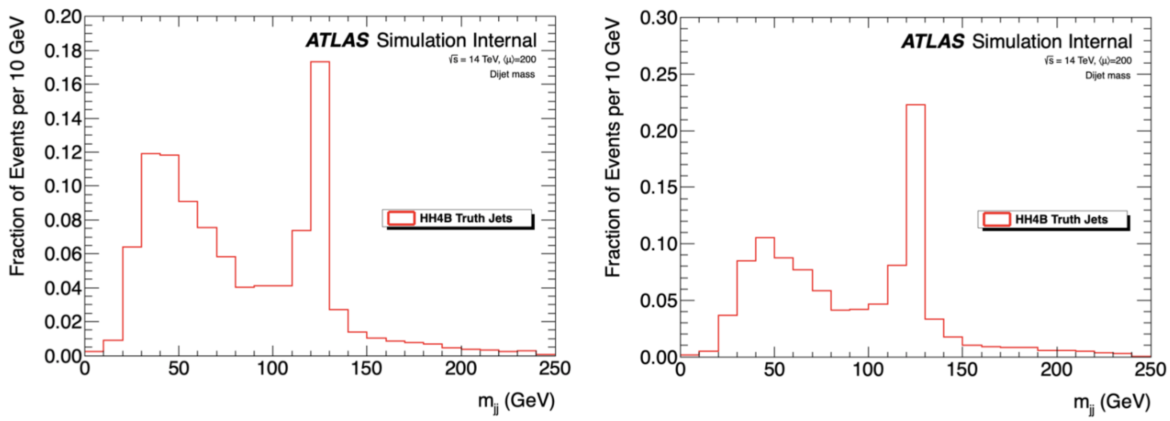

Figure 17: Normalized histograms of reconstructed Higgs masses for multiple truth events. On the left is the distribution with no cuts applied. A distinct peak occurs around 125 GeV, the Standard Model value of the Higgs mass. The background is caused by incorrect jet selections or pairings. On the right, a 40 GeV threshold was introduced during the jet selection, and if four jets could not pass the threshold, the event was discarded. This cut makes the peak more prominent compared to background, as expected.

consistent with this prediction, as demonstrated in Figure 17. Removing the peak from the plot would leave a background shape that quickly peaks and then gradually falls off as mass increases. This combinatoric shape is formed by pairing jets that do not share a common parent particle, in this case the Higgs boson. A peak forms when a significant amount of reconstructed parent particles have similar masses. The significant accumulation of data around a specific point is called a resonance.

Using the same reconstruction algorithm on 420, 422, and USCA jets from a simulated HH4b sample should also produce resonances and combinatoric background. If, instead, the sample was made of non-related jets, the distribution should be only background without any significant peaks. To test this, a JZ0W sample containing jets with minimum bias was used. The results in Figure 18 and 19 show the actual shapes for all three algorithms using HH4b and JZ0W data.

If there are resonances in the 420, 422, and USCA algorithm, they do not appear as distinctly as in the truth plot using the current reconstruction algorithm. This lack of peak is due to the fact that the jets are reconstructed from clusters and might have differed in position or energy depending on the efficiency of the algorithm.

To obtain more defined peaks, some criteria must be introduced that minimizes incorrect di-jet selections. Further analysis suggested that events where the six selected jets were at low p_{T} became bad candidates for reconstruction. Ensuring that at least four jets in the event had a high p_{T} made it much more likely that a bottom quark pair would be selected. For this, a 40 GeV threshold was introduced, and any event without four jets over this threshold was discarded. This criteria was effective in rejecting combinatoric background

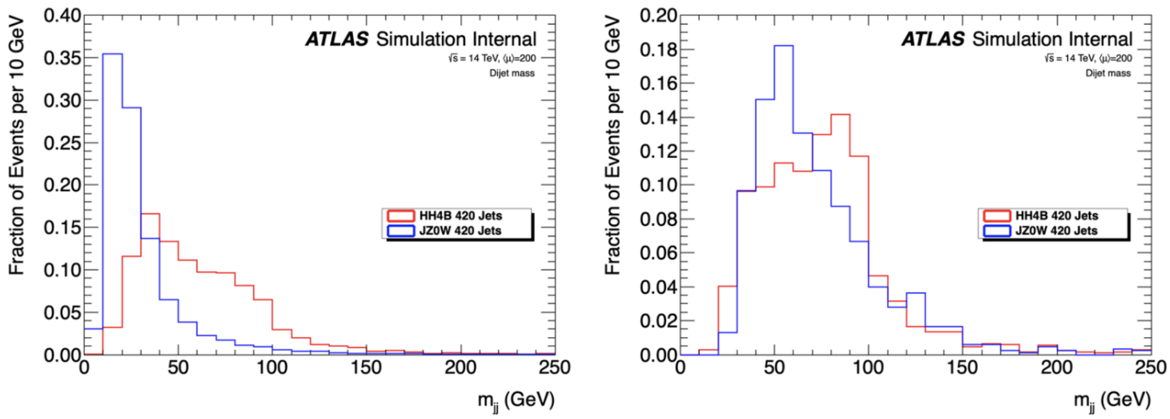

Figure 18: Reconstructed di-jet mass distribution using 420 jets. The left plot has no cuts introduced. Both HH4b and JZ0W samples create a combinatoric background shape, without any clearly defined peaks. Both USCA and 422 results are plotted to demonstrate the similar behavior of the two algorithms. The right plot was generated after a 4-jet 40 GeV threshold was introduced. The cut causes the two samples to create different shapes, with the HH4b sample containing a non-zero peak.

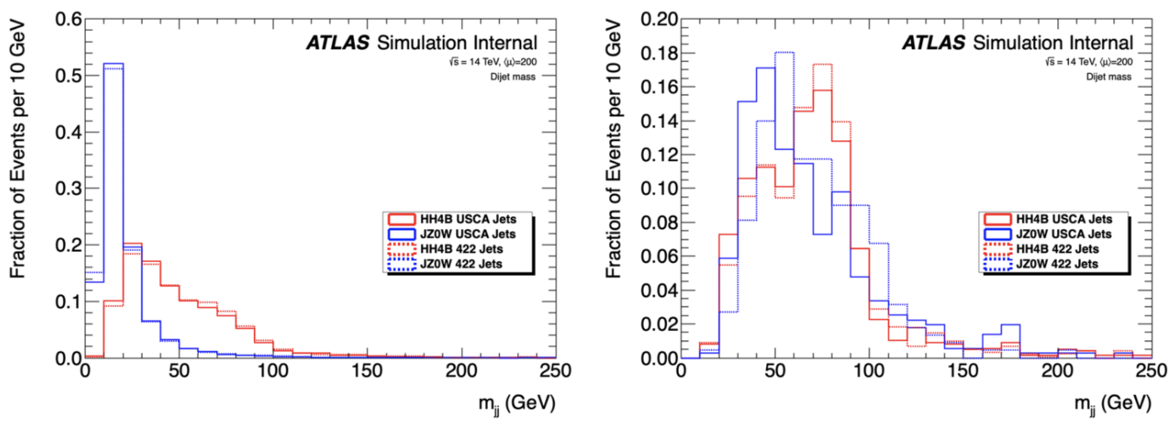

Figure 19: USCA and 422 di-jet mass reconstructions for HH4b and JZ0W sample. The left plot contains no significant peaks above background for either dataset. The right plot introduces a 4-jet 40 GeV threshold, Both USCA and 422 results are plotted to demonstrate the similar behavior of the two algorithms.

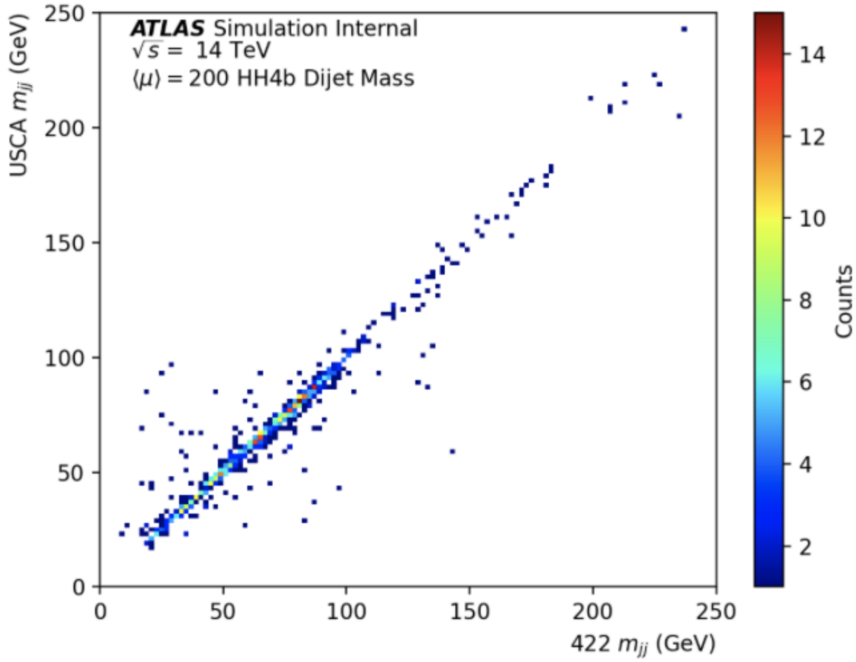

Figure 20: Histograms comparing the masses obtained by the reconstruction algorithm for the USCA and 422 jets on the HH4b dataset. A linear correlation near the center of the plot signifies agreement between the two algorithms.

while keeping a majority of resonant mass reconstructions. It also maintains a lack of peak in the JZ0W sample, suggesting the algorithm is not biased to produce a resonance.

To provide a more detailed comparison of the USCA and 422 algorithms, the same algorithm was run again with the 40 GeV cuts, and the resulting dijet masses from each algorithm were plotted on a two dimensional histogram. If the algorithms performed similarly, the resulting masses should be nearly identical. As seen in Figure 20, the two masses correlate well in the HH4b sample.

5 Conclusion/Outlook

The difficulty faced by modern accelerator physics is the high precision at which measurements have to be made. The models used to describe the universe are extremely accurate and comprehensive. Anomalies in these models are therefore infrequent and occur in extreme conditions, such as TeV-level energies. Violations of the Standard Model are expected to have relatively tiny cross-sections, so only by increasing the energy and rate of collisions can any evidence of them be found.

With every bunch crossing that occurs inside the ATLAS detector, a variety of collision and decay events take place. Most of these events are well-studied and do not need to be saved for further analysis. By increasing the energy and frequency of collisions, the detector will detect an overwhelming amount of unimportant events. The Global Trigger system is essential because it distinguishes the interesting events, such as Di-Higgs production, using various criteria. Events that do not pass any criteria are dumped, reserving memory space and reducing the frequency of data storage.

Since the Global Trigger is being introduced during the first run of the High Luminosity Large Hadron Collider, its ability to efficiently handle data is essential to the success of ATLAS. For this reason, the criteria on which it selects must be carefully fine tuned to maximize interesting data selection, while rejecting as many uninteresting events as possible. Furthermore, the incoming information, or trigger objects (TOBs), on which the criteria will be applied must be accurate.

One of these TOBs is the number and energies of jets detected in the event. To extract jet information from individual detector cells, a topological clustering algorithm must be employed. Testing such an algorithm on simulated data provides an indication of its effectiveness, especially data resembling events that are currently being searched for, like Di-Higgs production. Offline algorithms like the 422 and 420 algorithms are too slow to be implemented on the Global Trigger, but produce good reconstructions of jets from a collision. The results from faster algorithms like the USCA should be compared to results from the offline algorithms to demonstrate effectiveness. Being able to produce mass resonances using USCA-reconstructed jets from an HH4b sample is one example of how such a comparison can be done.

One of the primary motivators behind topological clustering in the Global Trigger is to identify HH4b events, which is why a simulated HH4b sample was used to test effectiveness. Figure 20 suggests the performance of the USCA and 422 algorithms are extremely similar for this sample, a valid indicator of the USCA algorithm's potential success in the actual trigger system. However, more studies on the difference between the two algorithms are necessary to set expectations for the Global Trigger.

By measuring tightly bound predictions of the Standard Model, experimentalists can discover new ways in which the model fails to explain physical processes. These failures help motivate theories that attempt to go beyond the Standard Model. By expanding our understanding of particle physics, physicists can create a more thorough explanation of how the universe was formed and why it exists in the way we perceive it today.

6 References

[1] S. Abachi et al. "Observation of the Top Quark". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[2] Ahmed Abokhalil. *The Higgs Mechanism and Higgs Boson: Unveiling the Symmetry of the Universe.* 2023. arXiv: 2306.01019 [hep-ph].

[3] A. Arbey and F. Mahmoudi. "Dark matter and the early Universe: A review". In: *Progress in Particle and Nuclear Physics* 119 (July 2021), p. 103865. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2021.103865. URL: http://dx.doi.org/10.1016/j.ppnp.2021.103865.

[4] ATLAS Collaboration. "A Particle Consistent with the Higgs Boson Observed with the ATLAS Detector at the Large Hadron Collider". In: *Science* 338 (2012), p. 1576. DOI: 10.1126/science.1232005.

[5] ATLAS Collaboration. *ATLAS LAr Calorimeter Phase-II Upgrade: Technical Design Report.* ATLAS-TDR-027; CERN-LHCC-2017-018. 2017. URL: https://cds.cern.ch/record/2285582.

[6] ATLAS Collaboration. *ATLAS TDAQ Phase-II Upgrade: Technical Design Report.* ATLAS-TDR-029; CERN-LHCC-2017-020. 2017. URL: https://cds.cern.ch/record/2285584.

[7] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[8] ATLAS Collaboration. "The ATLAS Simulation Infrastructure". In: *Eur. Phys. J. C* 70 (2010), p. 823. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568 [physics.ins-det].

[9] Sylvia Berryman. "Democritus". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta and Uri Nodelman. Spring 2023. Metaphysics Research Lab, Stanford University, 2023.

[10] "branching fraction". In: (2019). DOI: doi:10.1351/goldbook.B00725. URL: https://doi.org/10.1351/goldbook.B00725.

[11] Oliver Sim Brüning et al. *LHC Design Report.* CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[12] P Busch, T Heinonen, and P Lahti. "Heisenberg's uncertainty principle". In: *Physics Reports* 452.6 (Nov. 2007), 155–176. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2007.05.006. URL: http://dx.doi.org/10.1016/j.physrep.2007.05.006.

[13] Astrid Eichhorn, Aaron Held, and Christof Wetterich. “Predictive power of grand unification from quantum gravity”. In: *Journal of High Energy Physics* 2020.8 (Aug. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep08(2020)111. URL: http://dx.doi.org/10.1007/JHEP08(2020)111.

[14] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[15] X. Fan et al. “Measurement of the Electron Magnetic Moment”. In: *Phys. Rev. Lett.* 130 (7 Feb. 2023), p. 071801. DOI: 10.1103/PhysRevLett.130.071801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.130.071801.

[16] R. P. Feynman. “Space-Time Approach to Non-Relativistic Quantum Mechanics”. In: *Rev. Mod. Phys.* 20 (2 Apr. 1948), pp. 367–387. DOI: 10.1103/RevModPhys.20.367. URL: https://link.aps.org/doi/10.1103/RevModPhys.20.367.

[17] Apollinari G. et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1.* CERN Yellow Reports: Monographs. Geneva: CERN, 2017. DOI: 10.23731/CYRM-2017-004. URL: https://cds.cern.ch/record/2284929.

[18] Murray Gell-Mann. “The Eightfold Way: A Theory of strong interaction symmetry”. In: (Mar. 1961). DOI: 10.2172/4008239.

[19] M. Grazzini et al. “Higgs boson pair production at NNLO with top quark mass effects”. In: *Journal of High Energy Physics* 2018.5 (May 2018). ISSN: 1029-8479. DOI: 10.1007/jhep05(2018)059. URL: http://dx.doi.org/10.1007/JHEP05(2018)059.

[20] Werner Herr and B Muratori. “Concept of luminosity”. In: (2006). DOI: 10.5170/CERN-2006-002.361. URL: https://cds.cern.ch/record/941318.

[21] “Historical Introduction to the Elementary Particles”. In: *Introduction to Elementary Particles.* John Wiley & Sons, Ltd, 1987. Chap. 1, pp. 11–53. ISBN: 9783527618460. DOI: https://doi.org/10.1002/9783527618460.ch1. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527618460.ch1. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527618460.ch1.

[22] Takaaki Kajita. “Discovery of neutrino oscillations”. In: *Reports on Progress in Physics* 69.6 (May 2006), p. 1607. DOI: 10.1088/0034-4885/69/6/R01. URL: https://dx.doi.org/10.1088/0034-4885/69/6/R01.

[23] Miao Li et al. *Dark Energy: a Brief Review*. 2012. arXiv: 1209.0922 [astro-ph.CO].

[24] Gouranga C Nayak. *Matter-Antimatter Asymmetry Of The Universe and Baryon Formation From Non-Equilibrium Quarks and Gluons*. 2019. arXiv: 1909.05640 [physics.gen-ph].

[25] Emmy Noether. "Invariant variation problems". In: *Transport Theory and Statistical Physics* 1.3 (Jan. 1971), 186–207. ISSN: 1532-2424. DOI: 10.1080/00411457108231446. URL: http://dx.doi.org/10.1080/00411457108231446.

[26] RD Schaffer and Stanislaus Cornelius Maria Bentvelsen. *Definition of Offline Readout Identifiers for the ATLAS detector*. Tech. rep. revised version number 1 submitted on 2001-05-28 13:21:00. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/684167.

[27] Steven Weinberg. "The Search for Unity: Notes for a History of Quantum Field Theory". In: *Daedalus* 106.4 (1977), pp. 17–35. ISSN: 00115266. URL: http://www.jstor.org/stable/20024506 (visited on 11/16/2023).