SEARCHES FOR NEW PHYSICS USING UNSUPERVISED MACHINE

LEARNING FOR ANOMALY DETECTION AT THE ATLAS

DETECTOR AND THE DEVELOPMENT OF PARTICLE

IDENTIFICATION ALGORITHMS FOR THE HL-LHC

By

JACOB E. CROSBY

Bachelor of Science in Physics
Michigan State University
Lansing, Michigan
2019

Master of Science in Physics
Oklahoma State University
Stillwater, Oklahoma
2023

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2024

SEARCHES FOR NEW PHYSICS USING UNSUPERVISED MACHINE

LEARNING FOR ANOMALY DETECTION AT THE ATLAS

DETECTOR AND THE DEVELOPMENT OF PARTICLE

IDENTIFICATION ALGORITHMS FOR THE HL-LHC

Dissertation Approved:

Dr. Alexander Khanov

Dissertation Advisor

Dr. Mario Borunda

Dr. K. S. Babu

Dr. Cong Pu

ACKNOWLEDGMENTS

First off, I would like to thank my advisor, Dr. Alexander Khanov. Without his guidance and patience, my experience through graduate school could have been very different. Dr. Khanov accepted me as an REU student 6 years ago and introduced me to particle physics. The memorable time I had with him, and the team influenced my decision on choosing Oklahoma State University for graduate school which has completely changed my life. He has been a pivotal role, and I cannot thank him enough for being that. Dr. Flera Rizatdinova, thank you for your caring leadership. You're a vital piece that is required for the unity of our team.

I would also like to thank The High Energy Physics team at Oklahoma State University. You all have been my physics family and have truly been a large part of this special experience. To my fellow graduate friends and colleagues. You have all helped make OSU feel like home and I hope you find the warmth in others as I found in all of you. Sincere thanks to Dr. Alexander Khanov, Dr. K. S. Babu, Dr. Mario Borunda, and Dr. Cong Pu for accepting to be on my dissertation advisory committee.

To my friends overseas at CERN, thank you for making unforgettable memories with me and showing me how truly American I really am. Huge thanks to Luke, Chetna and Fabienne who acted with great urgency when I had a cooking accident making gnocchi, you all get two thumbs up. I would like to thank my colleagues in the Flavor Tagging team who helped reinforce my confidence and supplied guidance when sought.

I would like to thank the anomaly detection analysis team, Sergei Chekanov, Rui Zhang, and Wasikul Islam. Our synergy brought our analysis into fruition and helped pave the way for similar techniques. May you all remain innovative and efficient.

iii

Not many times in one's life do they get to experience a pandemic during their graduate degree. I would like to give credit to all those within OSU and the physics department for their patience and dedication to education, even through such chaos. We all certainly lived in interesting times.

Of course, my beloved family who have always been there with necessary support and have always had my back. I Love you all. Robyn Edwards, without this nexus of choices and people, we would never have met. I can never be thankful enough to the aligned stars that brought you into my life.

Lastly, I want to dedicate this thesis to my cousin and best friend who has passed before his time. Drew, I know you would be proud. This one is for you.

Name: JACOB E. CROSBY

Date of Degree: MAY, 2024

Title of Study: SEARCHES FOR NEW PHYSICS USING UNSUPERVISED MACHINE
                  LEARNING FOR ANOMALY DETECTION AT THE ATLAS DETECTOR
                  AND THE DEVELOPMENT OF PARTICLE IDENTIFICATION
                  ALGORITHMS FOR THE HL-LHC

Major Field: PHYSICS

Abstract: This document contains discussions on completed and ongoing projects that have developed over the past few years while working on the ATLAS detector located at CERN in Geneva, Switzerland. The first discussion will be on my qualification task for the ATLAS collaboration which is on the development and future plans of the DL1d tagger that is currently used as a baseline tagger for run 4 of the ATLAS detector. After, the discussion will transition to an analysis that applies a novel anomaly detection technique which uses a neural network architecture called the autoencoder. This neural network is then trained on 1% randomly selected events of run 2 data from the ATLAS detector. Once the model and anomalous regions are defined, the model is used to find phase spaces where events that contain physics beyond the standard model may occur. Statistical analysis is then applied to these phase spaces in order to find signatures of new physics. No significant signatures are found. Lastly, I will discuss an ongoing search for a new massive scalar X decaying into a new light scalar Y and the standard model Higgs boson H through the process X$\rightarrow$YH$\rightarrow$bbbb in the boosted topology.

TABLE OF CONTENTS

LIST OF TABLES

Figure Page

GLOSSARY

pp proton-proton. 26, 30, 34, 43, 53, 58, 64, 65, 78

AE autoencoder. 103, 107, 112–116, 121, 149, 150, 153, 155, 158, 181, 182

ALICE A Large Ion Collider Experiment. 26

AR Anomaly Region. 104, 118, 119, 121, 127, 128, 140, 143, 154, 155, 186, 188, 199, 210

ATLAS A Large Toroidal Aparatus. 3, 26, 29, 30, 33, 40, 43, 45, 51, 52, 56, 58, 65, 73, 74, 78–81, 84, 85, 93, 98–101, 103, 104, 114, 124, 148, 157, 168, 171, 174

BDT Boosted Decision Tree. 73–75, 81

BH BumpHunter. 126, 127, 140

BSM Beyond the Standard Model. 3, 9, 23, 30, 33, 78, 85, 103, 104, 106–109, 113, 116, 118, 121, 126, 140, 143, 144, 148, 149, 154, 155, 158, 181, 184, 186

CCA connected component analysis. 52

CERN Conseil Européen pour la Recherche Nucléaire", or European Council for Nuclear Research. 2, 26, 30

CKM Cabibbo-Kobayashi-Maskawa matrix. 8, 9, 12, 24

CL Confidence Level. 145, 146

CME center-of-mass energy. 27–29, 44, 90, 148

CHAPTER I

INTRODUCTION

As humans we are natural explorers, there are no challenges too daunting, no sought truth to be overwhelming. The thirst for answers about ourselves and our existence has driven the minds of men for centuries. This drive has started to uncover the possibility of almost limitless potential. Our ancestors have gazed into the face of the night sky for thousands of years, asking, dreaming, and fantasizing of stories from the beginning. As we've evolved and developed new tools and the footprints of society looming in the horizon, the same sky still gazed down, fueling our drive to muster forward. The pureness of our curiosity passed through generations started bearing fruit as new technologies were being developed. Logical reasoning and objective views of the the world in front of us stemmed hypotheses and models for its description and developed what we know today as science. Knowledge advanced so far that one day we split the atom, introducing a new era of science and physics, one of underlying anxiousness and fear but also excitement. These efforts towards understanding stretched and weaned its way into every facet of society giving us technology and access that previous civilizations would deem godlike. Here we find ourselves, delving even deeper into the field of physics. Searching for answers, patterns, and possibilities on every scale we can detect. The scale in which the following studies are performed is in the scale of high energy.

All the energy in the universe was produced in a single event called the Big Bang. In the moments after the Big Bang occurring, all of this energy was compacted in an extremely small volume as it rapidly expanded outwards. This environment created conditions that gave particles an enormous amount of energy, causing their relationships to be much different than what we see today. The parameters of this environment no longer exists anywhere in the universe except in particle accelerators here on Earth. These accelerators allow physicists to peer into this hostile

environment in order to gain a better understanding of the forces of nature and how they started. These accelerators can have a range of energies. Lower energies would be considered Nuclear Physics which is the study of the nuclei and the nucleus's atoms. The name for the atom was given by John Dalton in the 19th century after the Greek word *atomos* or "indivisible" [64]. Today we now know that an atom is not "indivisible" but is actually composed of smaller constituents called quarks. This scale can only be obtained with particle accelerators at higher energies. Thus, the study of High Energy Physics (HEP), or Particle Physics, is the study of the fundamental particles and the forces that connect them in the universe.

The leading institution of Particle Physics is the Conseil Européen pour la Recherche Nucléaire or the European Organization for Nuclear Research (CERN) located in Geneva, Switzerland. This institution is the world's largest particle accelerator with the Large Hadron Collider (LHC) being a 27 kilometer ring consisting of superconducting magnets along with a number of particle accelerating contraptions. During the 1950s and 1960s particle accelerators were designed for much higher energies than accelerators at that time. Within this new energy threshold, the renaissance of Particle Physics began. The majority of particles that exist aren't stable and are only produced via highly energetic events, so a perplexing amount of particles were observed for the first time in scattering experiments. These two decades were referred to as the "particle zoo". This term was no longer used in the early 1970s after the formulation of the Standard Model. This model is the foundation of Particle Physics and explains that these particles were a composition of a few much smaller and fundamental particles.

1.1 Outline

In chapter 2 we will discuss the Standard Model in much greater detail. Starting with its history and its obvious motivation. We will delve deeper into the theory, showing its beauty and discussing the importance of its creation. This will lead us into the physical signatures that the Standard Model explains and also further predicts which will lead us into the exciting discoveries it has

made decades later.

Chapter 3 follows up this discussion with explanations of tools developed in order to detect such signatures. This leads us into the birth of the Large Hadron Collider (LHC) and one its largest detectors, ATLAS. The inner workings of this detector will be explained along with how they paint a beautiful picture of the chaos that occurs inside. Not only is the current status of the ATLAS detector discussed but also its upgrade which is scheduled in 2029.

After the workings of the ATLAS detector is well established, chapter 4 will introduce object reconstruction, identification, and event simulation. From here we'll see how energy deposits within ATLAS leads to low level triggers and up to high level kinematic reconstruction. This will then lead us into the next section.

Once a full, reconstructed picture of the events that happen inside the ATLAS detector is well understood, chapter 5 will explain the complex and state of the art software tools developed and high level in order to create order from the chaos. This will transition into my work for my qualification task that helped create one of these tools using machine learning for the coming upgrade in run 4 for the ATLAS detector.

Now that the full picture of energy deposition in the ATLAS detector leading to high level object reconstruction is well understood, a discussion on finding Beyond the Standard Model (BSM) signatures can begin. Here, in chapter 6, a new and innovative technique is described that was created to find such signatures. This technique uses an agnostic and unsupervised machine learning approach to find anomalies within data of the ATLAS detector. This chapter covers from start to finish an entire analysis along with its findings.

Chapter 7 will finish the thesis with the application of this innovative approach into a non-agnostic BSM search. This analysis is still preliminary and its plots should not be taken as final results. However, it generates interesting discussion on adapting this anomaly detection technique. This analysis will be continued by another student.

Finally, chapter 8 will have the closing discussing and conclusions of this dissertation. Summing up three years of hard work and marking the beginning of another chapter in life.

CHAPTER II

THE STANDARD MODEL

```
"The effort to understand the universe is one
of the very few things that lifts human life
a little above the level of farce, and gives
       it some of the grace of tragedy."

        - Steven Weinberg, 1993
```

2.1   A Model of Leptons

Back in 1967 Steven Weinberg outlined the foundation of what would become the Standard Model (SM) in a three paged Physical Review letter titled *A Model of Leptons*. Here he stated "What could be more natural than to unite [leptons, photons, and their intermediate bosons] into a a multiplet of gauge fields?" [89]. This quote could not have been more of an understatement. The model that was predicted in this letter would later become the most successful theory ever conceived. The SM's articulate description of the universe predicted particles that were all later found including the latest particle, the Higgs boson.

This theoretical formulation was not conceived by Weinberg out of nowhere. It was an inevitable outcome from a composition of works. The first came from Yang Chen-Ning and Robert Mills, they provided an explanation for strong interactions via gauge theory in their 1954 paper [93]. Then came Chien-Shiung Wu demonstrating the non-conservation of parity in the weak interaction in 1957 [92]. Sheldon Lee Glashow proposed the bold $SU(2) \times U(1)$ model that showed the possibility of symmetry between electromagnetic and weak interactions which in turn predicted the Z boson [63].

Steven Weinberg was in close contact with Abdus Salam during the year of 1961. This led Abdus Salam and his long time collaborator John C. Ward to proposing a very similar model to Glashow's SU(2) × U(1) [87]. Though both of these models still required the masses for the W and Z bosons to be inserted by hand making the model non-renormalizable and thus non-physical. Lastly, to put all the puzzle pieces together, Peter Higgs came in and demonstrated spontaneous symmetry breaking via the Higgs mechanism [68]. Three years later in just three pages, Weinberg formulated the first iteration of the SM in 1967 [89].

2.2   The Standard Model

The SM of particle physics is one of the most successful theories that has been proposed in modern physics. It has predicted elementary and composite particles that scientists are still discovering to this day with the latest being the Higgs boson. It has withstood countless experimental checks as global scientific communities from all around the world have given billions towards this vast area of research. Creating one of the largest collaborations in the world with this model at its heart. The SM is a model of symmetries and how these symmetries break, creating elementary particles. The elementary particles predicted by the SM can be split into two groups as seen in Figure 1. There are the matter particles (purple and green) and the force particles (red and yellow). The matter particles can also be called fermions and are split into three generations. There are the quarks (up, down, strange, charm, bottom, top) and there 3 leptons (electron, muon and the tau lepton) and their corresponding neutrino (electron neutrino, muon neutrino and tau lepton neutrino). There are then the force carrying bosons along with the Higgs boson which is responsible for the Higgs mechanism and is discussed later in this chapter. How these elementary particles interact with one another is described by the force carrying particles. These particles are the photon (γ), gluon, W and Z bosons. Each of these force carrying particles mediate a corresponding force. All physical phenomena are interactions of forces which include four types; the electromagnetic force, gravitational force, the strong force and the weak force. As the SM was in its infancy, it started to reveal the possibility that these forces were combined as a single force at one point in time. The

current state of the SM is unable to relate the gravitational force to the other three forces, though this problem is well known and decades of careers are being put forth to solve this hole.

Now each force has its own responsibility. The gravitational force is what everyone interacts with everyday since it's the reason we simply don't float off into space. Gravity is well described on a macro scale through the beautiful theory of General Relativity which describes the interactions between macro objects such as planets and stars, though this is beyond the scope of this thesis and is well described in the field of astrophysics and astronomy. Currently gravity's interaction on the scale of the SM is so small, it's negligible, and has no relevance in understanding elementary particle interactions (at least to our current knowledge). The predicted force carrying particle is called the Graviton and it has eluded experiments since its formulation. The electromagnetic force is used to communicate between charged particles. This force is magnitudes stronger than the gravitational force and therefore has repercussions dictating the relationships on all scales. The force carrying particle is called the photon which is a massless boson. This elementary particle is a quantum of the electromagnetic field. Meaning, it's the bare minimum particle involved in interactions within this field. The strong force is the fundamental interaction that binds elementary particles, creating larger composite particles called hadrons. This force is mediated by the boson called a Gluon. Gluons are massless vector bosons and the binding of quarks are in accordance with Quantum Chromodynamics (QCD). These bosons carry color charge which adds a layer of complexity and is the reason gluons participate in this interaction. The strong force can also be called the nuclear force since it's the reason protons and neutrons (both hadrons) can be bound together composing nuclei of various sizes. This force (as one might deduce) is the strongest of all forces with the strength of 100 times of the electromagnetic force, 10^6 times stronger than the weak force and amazingly 10^{38} times stronger than the gravitational force. Lastly there's the weak force. The weak force is mediated by the W and Z bosons. The W boson mediates the transfer of electric charge and therefore can be either positive or negative (W^+, W^-) and are each other's antiparticle. Whereas the Z boson is electrically neutral (Z^0) and mediates the transfer of momentum, spin and energy. It's also its own antiparticle. These two bosons are the reasons particles are able to undergo

Standard Model of Elementary Particles

Figure 1: The Standard Model of Particle Physics. This displays all three generations of fermions in purple and green along with their antiparticle. On the right in red shows the force carrying bosons along with the Higgs boson shown in yellow. Corresponding table of values can be found in Table 1 on page 25.

radioactive decay. They both also participate in nuclear fission and nuclear fusion.

2.2.1 Theoretical Foundation

The SM represents a gauge theory that describes the strong, weak and electromagnetic interactions The SM Lagrangian $\mathscr{L}_{SM}$ can be broken down into two Lagrangians representing the strong interaction $\mathscr{L}_{QCD}$ and the electroweak (EW)interaction $\mathscr{L}_{EW}$. As in the name, the EW theory is the combination of the electromagnetic and weak interactions. The Lagrangian $\mathscr{L}_{SM}$ is invariant under local gauge transformations with the symmetry groups [84].

$$\underbrace{SU(3)_C}_{\text{QCD}} \otimes \underbrace{SU(2)_L \otimes U(1)_Y}_{\text{electroweak}} \tag{2.1}$$

The interactions between the strong, weak and electromagnetic forces are through the exchange of spin-1 gauge fields. For the strong field; 8 massless gluons. For the weak; 3 massive bosons, $W^{\pm}$ and Z^0. For the electromagnetic; 1 massless photon. The fermionic particles in Table 1 and Figure 1 can be represented in a 3-fold family structure:

$$
\begin{bmatrix} \nu_e & u \\ e^- & d' \end{bmatrix}, \qquad \begin{bmatrix} \nu_\mu & c \\ \mu^- & s' \end{bmatrix}, \qquad \begin{bmatrix} \nu_\tau & t \\ \tau^- & b' \end{bmatrix} \tag{2.2}
$$

where each quark appears in three different colors:

$$
\begin{bmatrix} \nu_l & q_u \\ l^- & q_d \end{bmatrix} \equiv \begin{pmatrix} \nu_l \\ l^- \end{pmatrix}_L, \quad \begin{pmatrix} q_u \\ q_d \end{pmatrix}_L, \quad l_R^-, \ q_{uR}, \ q_{dR} \tag{2.3}
$$

plus their corresponding antiparticles. Here we see that the left-handed fields are $SU(2)_L$ doublets, whereas the right-handed partners transform as $SU(2)_L$ singlets.

The vacuum-induced breakdown of gauge symmetry initiates Spontaneous Symmetry Breaking (SSB) (as discussed later in section 2.2.3) within the EW group, resulting in the emergence of the electromagnetic subgroup.

$$
SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \qquad \overrightarrow{SSB} \qquad SU(3)_C \otimes U(1)_{QED} \tag{2.4}
$$

The SSB mechanism is the cause of the masses of the weak gauge bosons ($W^{\pm}, Z^0$) and gives rise to the appearance of a physical scalar boson in the SM which can be seen as the yellow "Higgs" in Figure 1. This also gives rise to the fermion masses and their mixings. These mixings keep track of weak decays, i.e. one quark transitioning to another quark. These mixings were first formulated into a 6-quark model by Kobayashi and Maskawa by generalizing the Cabibbo matrix, creating the Cabibbo-Kobayashi-Maskawa (CKM) matrix. This matrix keeps track of weak decay rates in three

generations of quarks [78].

$$
\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}
\tag{2.2.1}
$$

Here we see on the left side are the weak interaction doublet partners of down-type quarks. On the right side is the CKM matrix along with a vector of mass eigenstates of down-type quarks. The CKM matrix states the probability of transitions between one quark flavor j to another quark flavor i. These transitions are proportional to $|V_{ij}|^2$.

$$
\begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} = \begin{bmatrix} 0.97373 \pm 0.00031 & 0.2243 \pm 0.0008 & 0.00382 \pm 0.00020 \\ 0.221 \pm 0.004 & 0.975 \pm 0.006 & 0.0408 \pm 0.0014 \\ 0.0086 \pm 0.0002 & 0.0415 \pm 0.0009 & 1.014 \pm 0.029 \end{bmatrix}
\tag{2.6}
$$

Here we see in Eq. 2.6 the most recent transition probabilities as stated by the Particle Data Group (PDG) [91]. Now, we expect unitary of the CKM matrix but if we check it, even in the first row, we see:

$$
|V_{ud}|^2 + |V_{us}|^2 + |V_{ub}|^2 = 0.9985 \pm 0.0007
\tag{2.7}
$$

The difference from the theoretical unitary value of 1 has a standard deviation of 2.2σ which this gives an exciting strong indication of physics Beyond the Standard Model or BSM.

9

2.2.2 Fundamental Interactions

The Electromagnetic Interaction

Quantum Electrodynamics (QED) is the foundational knowledge of electromagnetism. All charged particles communicate with each other through the electromagnetic force carrier boson called the Photon. QED is a consequence of the $U(1)$ symmetry. It is the separated force from the EW interactions and can be seen in Eq. 2.1 as $U(1)_{em}$. In order to obtain the QED Lagrangian, let's look at the dynamics of a free $1/2$ spin fermion.

$$\mathcal{L}_0 = \bar{\psi}(i\gamma^\mu - m)\psi \tag{2.8}$$

In Eq. 2.8 notation-wise, ψ is the Dirac Spinor, m is the mass and the Dirac matrices are denoted by $\gamma^\mu \cdot \bar{\psi} = \psi^\dagger \gamma^0$ which is also known as the Dirac adjoint. In the Lagrangian, the γ^μ are the Dirac 4×4 matrices and can be seen in Eq. 2.9.

$$\gamma^0 = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}, \quad \gamma^i = \begin{bmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{bmatrix}, \quad \gamma^5 = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix} \tag{2.9}$$

These are written in terms of the Pauli matrices which are shown in Eq. 2.10.

$$\sigma^1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2 = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}, \quad \sigma^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.10}$$

The Lagrangian in Eq. 2.8 is invariant under global gauge transformations, but is not invariant under local $U(1)$ transformations. In order to induce local invariance under transformations, as shown in Eq. 2.11, an additional vectorial gauge field that is massless is required.

$$\psi(x) \rightarrow \psi'(x) = e^{i\alpha(x)}\psi(x) \tag{2.11}$$

The gauge field that is required is denoted as $A_\mu(x)$ and is shown in eq 2.12. Additionally, the covariant derivative $D_\mu(x)$ is shown in Eq. 2.13.

$$A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\alpha(x) \tag{2.12}$$

$$D_{mu}(x) = \partial_{mu} - ieA_\mu(x) \tag{2.13}$$

Using these two fields, the field strength tensor can be expressed in Eq. 2.14.

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.14}$$

Now that local $U(1)$ symmetry is applied, Eq. 2.11 can be used along with Eq. 2.14 to express the QED Lagrangian. This is shown in Eq. 2.15.

$$\mathcal{L} = \bar{\psi}i\gamma^\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{2.15}$$

From here, the interaction term called the electromagnetic charge current density j^μ is added to Eq. 2.15. This equation is defined in Eq. 2.16.

$$j^\mu = \bar{\psi}\gamma^\mu\psi \tag{2.16}$$

putting both of these equations together we get Eq. 2.17.

$$\mathcal{L} = \underbrace{\bar{\psi}i\gamma^\mu - m)\psi}_{\text{free Lagrangian}} - \underbrace{ej^\mu A_\mu}_{\text{interaction term}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{kinetic term}} \tag{2.17}$$

Here, Eq. 2.17 is the QED Lagrangian. According to Noether's theorem which states that every differentiable symmetry of the action of a physical system with conservative forces has a corresponding conservation law [81]. For QED, this conserved quantity is the electromagnetic charge q. The QED Lagrangian thus shows the relationship between the photon field A_μ and the Dirac fields

11

$\boldsymbol{\psi}$ which emerged as a consequence of $\boldsymbol{U(1)}$ symmetry.

The Electroweak Interaction

The EW force is the combination of the electromagnetic interaction and the weak interaction. Both of these forces appear differently in low energies but they happen to combine at much higher energies. Thus, this theory models them as two different aspects of the same force. The predicted unification energy is on the order of 246 Gev or on a temperature scale of approximately $\mathbf{10^{15}K}$. This implies that the two forces coexisted at the start of the Big Bang and later diverged during the Quark Epoch, occurring approximately $\mathbf{10^{-12}}$ seconds after the inception of the Big Bang.

The mathematical formulation is found within the $\boldsymbol{SU(2)_L}$ symmetry group in Eq. 2.1. Weak isospin and weak hypercharge are quantum numbers relating the electrically charged part of the weak interaction and are labeled $\boldsymbol{T_i}$ and $\boldsymbol{Y_W}$ respectively. These are known as generators of $\boldsymbol{SU(2)}$ and $\boldsymbol{U(1)}$ and give rise to the gauge bosons that mediate this force. Once SSB occurs, these bosons are seen in the SM as the $\boldsymbol{W^\pm}$, $\boldsymbol{Z^0}$, and the photon. The conserved quantity within the electroweak force is the third component of the generator $\boldsymbol{T_i}$, quantum number $\boldsymbol{T_3}$. However, interaction with the Higgs field do not conserve weak isospin $\boldsymbol{T_3}$ and thus causing fermion mixings as seen in the CKM matrix in 2.6. Though, there are specific combinations of them that do not interact with the Higgs field and therefore are conserved, this happens to be the electric charge q. The combinations that give rise to q are given by Eq. 2.18.

$$Q = T_3 + \frac{1}{2}Y_W \tag{2.18}$$

The EW force is *chiral*, which requires treating both components of the fermionic fields $\boldsymbol{\psi}$ separately. Left-handed fermions have weak isospin of $\boldsymbol{T_3 = \pm 1/2}$ and are represented by doublets $\boldsymbol{\psi_L}$. Whereas right-handed fermions have weak isospin of $\boldsymbol{T_3 = 0}$ and are represented by singlets $\boldsymbol{\psi_R}$. Both components behave differently under $\boldsymbol{SU(2)_L}$ and $\boldsymbol{U(1)_\gamma}$ local transformations. A quick formulation of the EW Lagrangian can be formulated by introducing the covariant derivative acting

on both the left-handed fermionic field (Eq. 2.19) and the right-handed fermionic field (Eq. 2.20).

$$D_\mu \psi_L = (\partial_\mu + ig\frac{\sigma_i}{2}W_\mu^i + ig'\frac{\gamma}{2}B_\mu^i)\psi_L \tag{2.19}$$

$$D_\mu \psi_R = (\partial_\mu + ig'\frac{\gamma}{2}B_\mu^i)\psi_R \tag{2.20}$$

Here we see the Pauli matrices σ that are shown in Eq. 2.10, g and g' are coupling constants for the W^i and B_μ boson field strength tensors which are shown in Eq. 2.21 and Eq. 2.22.

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.21}$$

$$W_{\mu\nu}^i = \partial_\mu W_i^\nu - \partial_\nu W_i^\mu - \epsilon_{ijk}W_j^\mu G_k^\nu \tag{2.22}$$

With these, the electroweak Lagrangian can be assembled as shown in Eq. 2.23 where the sum over j covers the L doublet and two R singlets.

$$\mathscr{L}_{EW} = \sum_{j=1}^{3} \bar{\psi}_j[i\gamma^\mu D_\mu]\psi_j - \frac{1}{4}W_i^{\mu\nu}W_{\mu\nu}^i - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{2.23}$$

Both, $W^\pm$ bosons arise from linear combinations between W_1 and W_2 as shown in Eq. 2.24.

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \tag{2.24}$$

Finally, the application of a rotation by an angle θ_W allows the restoration of the massless vector field A that's associated with the photon and the massive weak neutral Z^0. Here, θ_W represents the *weak mixing angle*. This rotation introduces the mismatch between the $W^\pm$ bosons and the Z^0 boson. Through SSB, the two bosons (photon and Z^0) become physical with different masses given by Eq. 2.25.

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} cos\theta_W & sin\theta_W \\ -sin\theta_W & cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \tag{2.25}$$

where the offset of the Z^0 mass from the $W^\pm$ bosons can be given by Eq. 2.26.

$$m_Z = \frac{m_W}{cos\theta_W} \tag{2.26}$$

A pictorial representation of this *weak mixing angle* showing the pattern of weak isospin, T_3, and weak hypercharge, Y_W, of the known elementary particles can be seen in Figure 2.
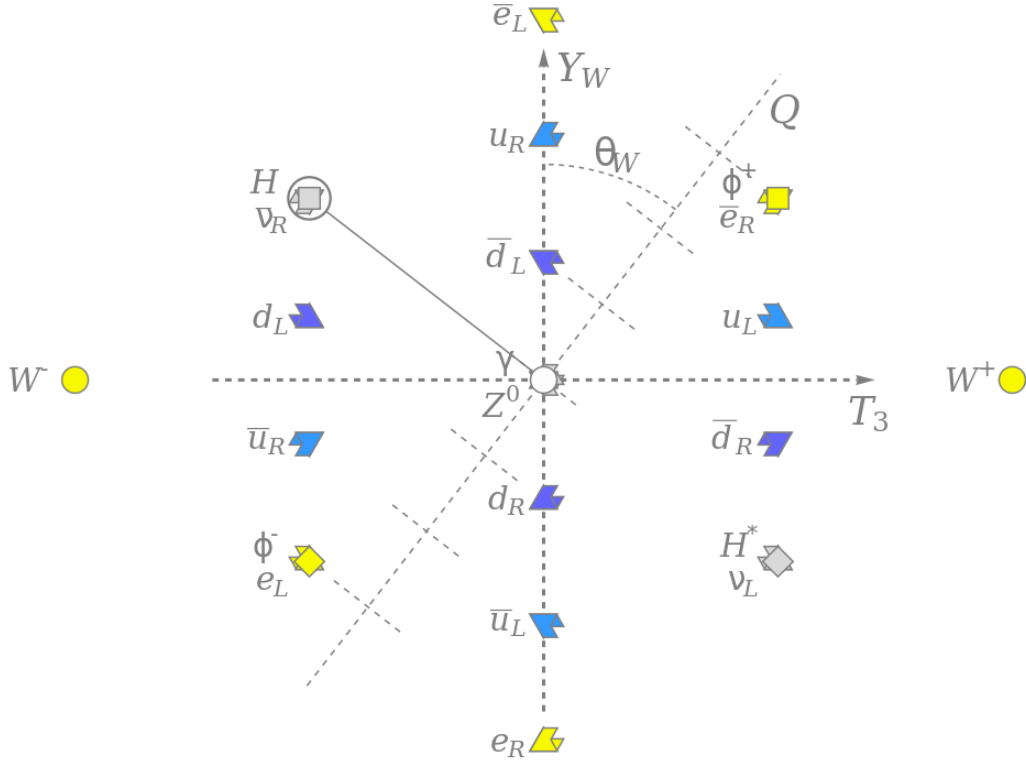
Figure 2: The pattern of weak isospin, T_3, and weak hypercharge, Y_W, of the known elementary particles. The electric charge is shown as Q along the *weak mixing angle*. The neutral Higgs field is seen being circled, this field breaks the electroweak symmetry and interacts with other particles, giving them mass.

The Strong Interaction

The strong interaction, or the strong force, is the fundamental force that confines quarks, forming hadrons. It is also the reason why nuclei are able to bind inside the nucleus forming atomic nuclei. To model the strong interaction, Quantum Chromodynamics, i.e. QCD, is used. This theory was founded on Yang-Mills theory [93], which extends QED, adding the non-abelian symmetry group $SU(3)$. Quarks interact with gluons and each other by way of a type of charge called the color charge. Unlike the electromagnetic charge, color charges comes with three types ($\pm$red, $\pm$green, $\pm$blue) which dictates the rules of behavior. Due to the strength of this force, when hadrons collide with high energy particles, jets of massive particles are produced instead of emitting their constituents as separate particles.

In the symmetry group $SU(3)$ Lie algebra, there are eight generators, therefore just as many gauge fields need to be introduced. The associated fermionic fields are represented by three-dimensional vectors, i.e. triplets, where each component represents color charges. Matrices called the Gell-Mann matrices t_a are used to represent the generators, having the commutation relationship: $[t_a, t_b] = if_{abc}t_c$. Under a transformation $U(x)$, the quark field ψ behaves as stated in Eq. 2.27.

$$\psi(x) \rightarrow \psi'(x) = U(x)\psi = e^{i\alpha_a(x)t_a}\psi \tag{2.27}$$

Similarly to the QED's case, new fields are introduced: the gluon field G_μ^a and a covariant derivative, as seen in Eq. 2.28.

$$D_\mu = \partial_\mu + ig_s t_a G_\mu^a \tag{2.28}$$

The gluon field behavior has requirements under the transformation $U(x)$. These requirements are shown in Eq. 2.29.

$$G_\mu^a \rightarrow G_\mu'^a = U(x)G_\mu^a t^a U^\dagger(x) + \frac{i}{g_s}(d_\mu U(x))U^\dagger(x) \tag{2.29}$$

The field strength tensor for the gluon field G_μ^a that follows the requirements as stated in Eq. 2.29 is shown in Eq. 2.30.

$$G_a^{\mu\nu} = \partial_\mu G_a^\nu - \partial_\nu G_a^\mu - g_s f_{abc} G_b^\mu G_c^\nu \tag{2.30}$$

Lastly, the QCD Lagrangian density, using the previously shown equations, can be derived and is shown in Eq. 2.31.

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a \tag{2.31}$$

Here, we see a noteworthy aspect of the QCD Lagrangian which is the self interaction between gluons due to having color charge.

2.2.3 Spontaneous Symmetry Breaking

The gauge symmetries of the SM ($SU(3)_C$, $SU(2)_L$, $SU(1)_\gamma$) have proven to be successful in predicting interactions between its particles. But particles have masses (besides the photon)! So far, there has been no references to any mass terms within these symmetries. Another mechanism was needed in order to allow generation of masses within these gauge theories. This mechanism is called spontaneous symmetry breaking (SSB).

Suppose there is a scalar field ϕ with a corresponding Lagrangian.

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) + e^{-(\alpha \phi)^2} \tag{2.32}$$

There is no obvious mass term within this Lagrangian, but if the exponential is expanded, we get:

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) + 1 - \alpha^2 \phi^2 + \frac{1}{2}\alpha^4 \phi^4 - \frac{1}{6}\alpha^6 \phi^6 + \dots \tag{2.33}$$

Here, the 1 is irrelevant, but the second term is close to the known mass term within the Klein-Gordon equation, with $\alpha^2 = \frac{1}{2}(mc/\hbar)^2$, while the higher order terms correspond to coupling terms. The Lagrangian describes a particle of mass:

$$m = \sqrt{2}\alpha \hbar/c \tag{2.34}$$

Now that the hidden mass term can be found within a Lagrangian for a field, suppose there's a Lagrangian that takes the form:

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) + \frac{1}{2}\mu^2 \phi^2 - \frac{1}{4}\lambda^2 \phi^4 \tag{2.35}$$

Here the second term looks like the mass term as previously discussed be we see that the sign is flipped and therefore would be imaginary. In order to properly understand this, Feynman calculus must be used which treats this more like a perturbation procedure which is started from the ground

state, or vacuum. We must represent this Lagrangian classically, as seen in 2.36, in order to obtain

the minimum.

$$\mathcal{L} = T - U \tag{2.36}$$

We obtain the minimum:

$$U(\phi) = -\frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda^2\phi^4 \tag{2.37}$$

We see that the minimum occurs at:

$$\phi = \pm\mu/\lambda \tag{2.38}$$

A new variable, η, must be introduced which represents a perturbation around this ground state.

$$\eta \equiv \phi \pm \frac{\mu}{\lambda} \tag{2.39}$$

Now, to rewrite the Lagrangian in Eq. 2.35 in terms of η.

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \mu^2\eta^2 \pm \mu\lambda\eta^3 - \frac{1}{4}\lambda^2\eta^4 + \frac{1}{4}(\mu^2/\lambda)^2 \tag{2.40}$$

The second term is now the correct sign for the mass term! The mass of the particle from the

Lagrangian is:

$$m = \sqrt{2}\mu\hbar/c \tag{2.41}$$

Where the third and fourth term are corresponding coupling terms. Figure 3 shows the shape of

the potential $U(\phi)$ and its minima. This example illustrates SSB, though it may not be too obvious.

The Lagrangian in Eq. 2.40 is the same as Eq. 2.35 except for one thing, its symmetry is broken.

The Lagrangian in Eq. 2.35 is *even* in ϕ, as in, it's invariant ($\phi \rightarrow -\phi$). Whereas the Lagrangian

in 2.40 is not, thus SSB. This happens due to the fact that the chosen state, i.e. the ground state,

does not share this symmetry. However, the collection of all states does share this symmetry. This

example shows a broken *discrete* symmetry, i.e left or the right side in Figure 3. To make this more
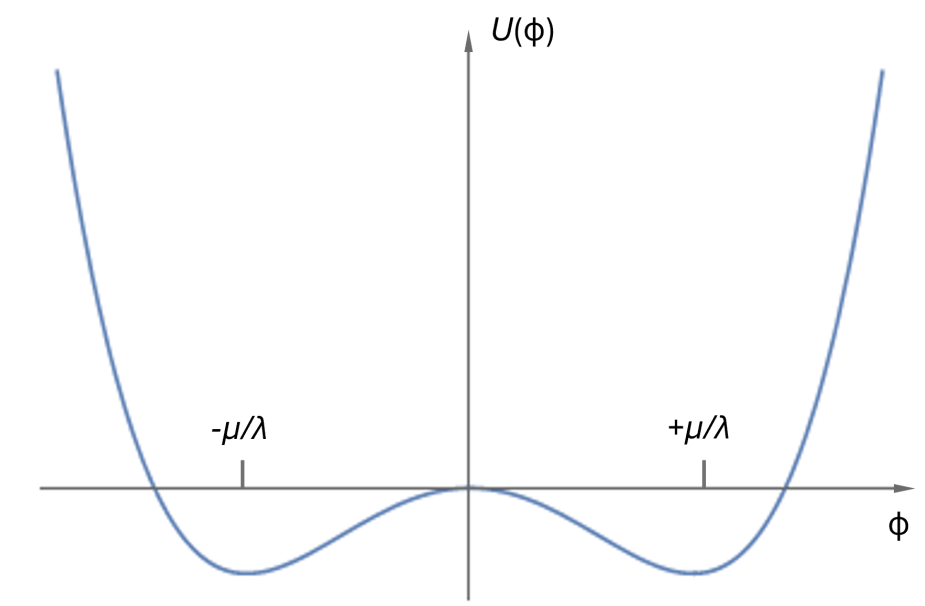
Figure 3: Graph of $U(\boldsymbol{\phi})$ in Eq. 2.37

physical, let's consider a Lagrangian containing continuous symmetries that demonstrates SSB.

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \phi_1)(\partial^\mu \phi_1) + \frac{1}{2}(\partial_\mu \phi_2)(\partial^\mu \phi_2) + \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) - \frac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 \tag{2.42}$$

Now this has two fields, $\boldsymbol{\phi_1}$ and $\boldsymbol{\phi_2}$, and only contains sum of squares, therefore it is invariant under *rotations* in $\boldsymbol{\phi_1}$, $\boldsymbol{\phi_2}$ space. The potential $\boldsymbol{U(\phi_1, \phi_2)}$ is then:

$$U(\boldsymbol{\phi_1}, \boldsymbol{\phi_2}) = -\frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) + \frac{1}{4}\lambda^2(\phi_1^2 + \phi_2^2)^2 \tag{2.43}$$

and the minima lie on the circle of radius $\boldsymbol{\mu / \lambda}$:

$$\phi_1^2 + \phi_2^2 = \mu^2 / \lambda^2 \tag{2.44}$$

Now in order to apply Feynman calculus as we did previously, we must expand around a particular ground state or vacuum.

$$\boldsymbol{\phi_1} = \boldsymbol{\mu / \lambda} : \boldsymbol{\phi_2} = \boldsymbol{0} \tag{2.45}$$

19

Following the steps shown above, we introduce new fields that represent fluctuations around the
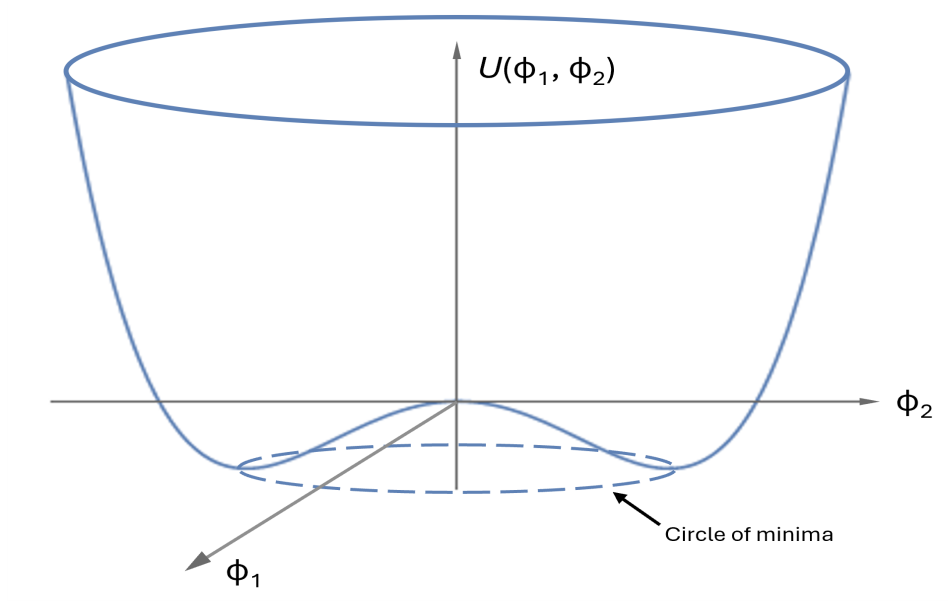
Figure 4: Graph of $U(\phi_1, \phi_2)$ in Eq. 2.43

vacuum.

$$\eta \equiv \phi_1 - \mu/\lambda; \qquad v \equiv \phi_2 \tag{2.46}$$

Rewriting the Lagrangian from Eq. 2.42 using these new fields:

$$\begin{aligned}
\mathcal{L} =& [\frac{1}{2}(\partial_\mu \eta \partial^\mu \eta) - \mu^2 \eta^2] + [\frac{1}{2}(\partial_\mu v \partial^\mu v)] - \\
& [\mu\lambda(\eta^3 + \eta v^2) + \frac{\lambda^2}{4}(\eta^4 + v^4 + 2\eta^2 v^2)] + \frac{\mu^4}{4\lambda^2}
\end{aligned} \tag{2.47}$$

The first term is the free Klein-Gordon Lagrangian for the field η, carrying a mass of:

$$m_\eta = \sqrt{2}\mu\hbar/c \tag{2.48}$$

The second term is a free Lagrangian for the field v, which is massless:

$$m_v = 0 \tag{2.49}$$

The third and fourth term show five different couplings. An important phenomena that occurs here

20

is that one of the two fields is *massless*. This is to be expect since SSB of a continuous symmetry is always accompanied by one or more massless scalar (spin-0) particles, these are called Goldstone Bosons. At first glance this may seem like an issue since there are no massless scalar bosons within the SM. At this point, another mechanism is needed to extract masses. This mechanism is called the Higgs Mechanism [68].

2.2.4 The Higgs Mechanism

This mechanism has been one of the most influential additions to the SM and is embodied by the Higgs boson. The combination of SSB and local gauge invariance has given way to the mechanism that has since accurately predicted masses for gauge fields, explaining the massive bosons $W^{\pm}$ and Z^0.

To start, the Lagrangian in Eq. 2.35 must be written with the combination of two fields, ϕ_1 and ϕ_2 into a complex field:

$$\phi \equiv \phi_1 + i\phi_2 \tag{2.50}$$

It is written like this so that its minima lie on a circle as seen in Figure 4:

$$\phi^*\phi = \phi_1^2 + \phi_2^2 \tag{2.51}$$

The Lagrangian in Eq. 2.35 can be rewritten as:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^*(\partial^\mu \phi) + \frac{1}{2}\mu^2(\phi^*\phi) - \frac{1}{4}\lambda^2(\phi^*\phi)^2 \tag{2.52}$$

Now the trick here is to not only introduce a massless gauge field A^μ, but the partial derivatives are replaced with *covariant* derivatives.

$$\mathfrak{D} = \partial_\mu + i\frac{q}{\hbar c}A_\mu \tag{2.2.2}$$

21

Rewriting Eq. 2.52:

$$\mathcal{L} = \frac{1}{2}[(\partial_\mu - i\frac{q}{\hbar c}A_\mu)\phi]^*[(\partial^\mu + i\frac{q}{\hbar c}A^\mu)\phi] +$$
$$\frac{1}{2}\mu^2(\phi^*\phi) - \frac{1}{4}\lambda^2(\phi^*\phi)^2 - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} \tag{2.53}$$

Now to define two new fields in order to fluctuate about the ground state.

$$\eta = \phi_1 - \mu/\lambda; \qquad \xi = \phi_2 \tag{2.54}$$

Once these fields are substituted and the Lagrangian is expanded, it will output a mass term but will also include a bilinear interaction and a massless Goldstone boson. To do away with the useless nonsense, a clever trick can be implemented. The complex field can be rewritten in terms of its real and imaginary parts as shown:

$$\phi \rightarrow \phi' = (cos\theta + isin\theta)(\phi_1 + i\phi_2)$$
$$\phi' = (\phi_1 cos\theta - \phi_2 sin\theta) + i(\phi_1 sin\theta + \phi_2 cos\theta) \tag{2.55}$$

If θ is picked as:

$$\theta = -\tan^{-1}(\phi_2/\phi_1) \tag{2.56}$$

It will ensure that ϕ' is real and that the second field, ξ, is dropped. Now the Lagrangian in Eq. 2.53 can be expanded out to get the final result.

$$\mathcal{L} = [\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \mu^2\eta^2] + [-\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}(\frac{q}{\hbar c}(\frac{\mu}{\lambda})^2)A_\mu A^\mu] +$$
$$[\frac{\mu}{\lambda}(\frac{q}{\hbar c})^2\eta(A_\mu A^\mu) + \frac{1}{2}(\frac{q}{\hbar c})^2\eta^2(A_\mu A^\mu) - \lambda\mu\eta^3 - \frac{1}{4}\lambda^2\eta^4] + (\frac{\mu^2}{2\lambda})^2 \tag{2.57}$$

The top two terms within this Lagrangian shows a massive scalar η on the left, which happens to be the Higgs boson, and a massive gauge field A^μ on the right. While the other terms refer to unique interactions. Here we see that through the extraordinary combination of SSB and local

gauge invariance, the SM is able to generate masses for gauge fields.

2.3 Beyond the Standard Model

Despite the immense success of the SM with its ability to explain fundamental interactions and predict particle masses, it is still an incomplete model. There are quite a few natural phenomena that the SM does not explain. Though, it is worthy noting that there is currently no experimental result that contradicts the SM up to 5σ [75]. Physics that is not adequately explained by the SM is called Beyond the Standard Model or BSM.

2.3.1 Unexplained Phenomena

Some fundamental physical phenomena not explained by the SM are:

Gravity This fundamental force is not explained within the SM. There have been theories that have tried to explain it by adding its own particle called the *graviton* which is yet to be discovered. Also, the best theory to explain gravity called *General Relativity*, which was formulated by Albert Einstein, is incompatible with the SM.

Neutrinos Mass Neutrinos do not have mass within the SM, but experimental data and astronomical observations have shown that neutrinos oscillate between flavors, i.e. lepton family number, which is a process requiring them to have non-zero mass [2]. The mechanism in which allows neutrinos to have mass within the SM has yet to be discovered.

Dark Matter Cosmological observations have shown hints of the existence of dark matter. If dark matter does exist, the ordinary matter that the SM explains only makes up 5% of the universe whereas dark matter would make up 26%. Dark matter is predicted to behave like ordinary matter but it interacts with SM fields weakly. Currently, there are no fundamental particles that are well explained and there have been no experimental evidence of such.

Dark Energy Just as dark matter, dark energy has yet to be discovered experimentally but is theoretically predicted within *General Relativity*. Dark energy theoretically makes up 69% of

23

the universe. This energy is considered the energy density of the vacuum, using the SM to determine this energy density, the calculated value is mismatched by 120 magnitudes.

Matter and Anti-matter Asymmetry In Cosmological observations, it is seen that the universe has a vastly disproportionate amount of matter to anti-matter. The so-called Sakharov-conditions [86] state how this asymmetry may occur through the violation of Charge-Parity (CP)-symmetry. CP-violation can be observed through meson mixing [20] and is related to the CP-violating phase in the CKM-matrix. The amount of violation is insufficient to explain the large discrepancy of matter to anti-matter.

The Standard Model						
Category	Particle Name	Symbol	Mass	Spin	Electric Charge	Weak Isospin
Quarks	up	u	2.2 MeV	1/2	+2/3	+1/2
	down	d	4.7 MeV	1/2	-1/3	-1/2
	charm	c	1.3 GeV	1/2	+2/3	+1/2
	strange	s	93.4 MeV	1/2	-1/3	-1/2
	bottom	s	4.2 GeV	1/2	+2/3	+1/2
	top	t	173.2 GeV	1/2	-1/3	-1/2
Leptons	electron	e	0.511 MeV	1/2	-1	-1/2
	muon	μ	105.7 MeV	1/2	-1	-1/2
	tau	τ	1776.9 MeV	1/2	-1	-1/2
	electron neutrino	ν_e	0	1/2	0	+1/2
	muon neutrino	ν_μ	0	1/2	0	+1/2
	tau neutrino	ν_τ	0	1/2	0	+1/2
Gauge Bosons	gluon	g	0	1	0	
	photon	γ	0	1	0	0
	Z boson	Z	91.2 GeV	1	0	0
	W boson	W$\pm$	80.4 GeV	1	± 1	$\pm 1/2$
Higgs	Higgs boson	H	125.2 GeV	0	0	0

Table 1: The Standard Model (excluding antiparticles) particles and their relevant information [91].

CHAPTER III

THE ATLAS DETECTOR AT THE LHC

In order to study the SM and all of its parameters, the experiments require an immense amount of energy that is close to the levels of the Big Bang. This is no simple feat to say the least. Located at the facility of CERN is the world's most powerful particle accelerator called the Large Hadron Collider (LHC) [59]. Data from proton-proton (*pp*) collisions occurring at the LHC are recorded by the A Large Toroidal Apparatus (ATLAS) detector [29]. This chapter provides an overview of the LHC, the ATLAS detector and the detector's next upgrade.

3.1   The Large Hadron collider

The LHC is the world's largest machine spanning a 27 kilometer long circular tunnel buried 100 meters underground located at CERN near the border of France and Switzerland. This massive human feat accelerates protons (sometimes other hadrons) close to the speed of light in two beam pipes in opposing directions around its circular tunnel, giving them energy levels that closely resemble the Big Bang. There are four crossing points along the circumference where some of the world's largest particle detectors are located in order to collect collision data from the beam crossings. These are the ATLAS detector [29], the Compact Muon Solenoid CMS detector [32], the Large Hadron Collider beauty experiment (LHCb) [33], and the A Large Ion Collider Experiment (ALICE) [28]. Currently, there are in total 9 detectors located at the LHC. Billions of collisions occur every second within these beam crossings, creating an immense amount of stored data. The CERN Data Centers store more than 30 petabytes of data per year. The LHC succeeds this by using superconducting radio frequency (RF) cavities for particle acceleration and 1232 superconducting NbTi dipole magnets for controlling the particles circular path that operate at temperatures lower

LHC Specs as of Run 2	
Quantity	Number
circumference	26.658 km
Number of Magnets	9593
Number of Main Dipoles	1232
Dipole Operating Temperature	1.9K (-271.3°C)
Number of Main Quadrupoles	392
Number of RF Cavities	8 per beam
Nominal Energy (protons)	6.5 TeV
Nominal Energy (ions)	2.56 TeV/u (energy per nucleon)
Nominal Center-of-Mass Energy	13 TeV
No. of bunches per proton beam	2808
No. of protons per bunch	1.2×10^{11}
No. of turns per second	11245
No. of collisions per second	1 billion

Table 2: Specs of the Large Hadron Collider as of Run 2 [59].

than 2K with magnetic field strengths of up to 8.33T. A number of specs can be found about the LHC in table 2.

The LHC complex incorporates several smaller accelerators and subsystems to achieve the high energies used for state-of-the-art physics. Protons gain energies from smaller accelerators that lead up to the main ring of the LHC as seen in Figure 5.

Initially, the protons are produced by ionizing hydrogen atoms and then are fed into the LINAC 2 which raises their kinetic energy to 50 MeV [59]. Subsequently, the Proton Synchrotron Booster up to energies 1.4 GeV, the Proton Synchrotron to 25 GeV then the Super Proton Synchrotron up to 450 GeV. These are then injected into the LHC using the SPS where the protons are then accelerated to their final stages. Protons and heavy ions are accelerated in groups of particles, or bunches [59]. A proton bunch contains approximately 1.2×10^{11} protons at center-of-mass energy (CME) of 13 TeV as of Run 2 of the LHC. These proton bunches are 1.06 ns long. Up to 2808 proton bunches are circulating at one time during peak luminosity and are separated by at least 24.95 ns, distributed in about 3654 positions.

The CERN accelerator complex
Complexe des accélérateurs du CERN

LHC - Large Hadron Collider // SPS - Super Proton Synchrotron // PS - Proton Synchrotron // AD - Antiproton Decelerator // CLEAR - CERN Linear Electron Accelerator for Research // AWAKE - Advanced WAKefield Experiment // ISOLDE - Isotope Separator OnLine // REX/HIE - Radioactive EXperiment/High Intensity and Energy ISOLDE // LEIR - Low Energy Ion Ring // LINAC - LINear ACcelerator // n-ToF - Neutrons Time Of Flight // HiRadMat - High-Radiation to Materials // CHARM - Cern High energy AcceleRator Mixed field facility // IRRAD - proton IRRADiation facility // GIF++ - Gamma Irradiation Facility // CENF - CErn Neutrino platForm

Figure 5: Schematic of the Large Hadron Collider [59] showing all the accelerators and their relative positions

3.1.1 Cross Section, Luminosity and Pile-Up

All physics processes have a calculated rate in which they may occur. The probability that two protons will collide and interact a certain way is called the cross section $\sigma_X(\sqrt{s})$ which is a function of the CME $\sqrt{s}$. Particles in which have larger cross sections are more likely to occur. The calculated rate is based on the number N_X of collisions yielding some final state X. This rate R_X of production can be calculated by Eq. 3.1:

$$R_X = \frac{dN_X}{dt} = L \cdot \sigma_X(\sqrt{s}) \tag{3.1}$$

Where L is the integrated luminosity. The kinematics of any process is heavily dependent on the CME $\sqrt{s}$. The LHC has undertaken a few campaigns in its lifetime. What is considered Run 1 had a CME of $\sqrt{s} = 7$ TeV in 2010 and in 2011, while in 2012 it was $\sqrt{s} = 8$ TeV [56]. Run 2 obviously proceeded after once the LHC underwent a few upgrades. The campaign of Run 2 took place between the years of 2015 to 2018 with a CME of $\sqrt{s} = 13$ TeV [90], corresponding to a kinetic energy of 6.5 TeV per proton beam. Currently, as of 2022, the LHC has started Run 3 which has an increased CME of $\sqrt{s} = 13.6$ TeV. There are plans of having a technical upgrade of the LHC for Run 4 in 2029, making it the High Luminosity Large Hadron Collider (HL-LHC). This upgrade aims to have a CME of $\sqrt{s} = 14$ TeV. Studies within this thesis contain preliminary development of machine learning tools using simulated HL-LHC data along with an extensive physics search using data taken during Run 2.

Luminosity is a significant metric which is crucial in finding very rare physics processes. It is constructed of beam parameters: the number of proton bunches N_b, the number of protons per bunch, N_1 and N_2, (there are two since two bunches are colliding), the revolution frequency f_r of the bunches, and lastly the overlap in both x and y directions, σ_x and σ_y [67]. Giving Eq. 3.2.

$$\implies \mathcal{L} = \frac{N_b N_1 N_2 f_r}{4\pi\sigma_x\sigma_y} \tag{3.2}$$

This equation shows two Gaussian beam bunches colliding head-on. The LHC is designed to have a peak instantaneous luminosity of 1.0×10^{34} cm^{-2}s^{-1} which Run 2 was able to surpass. In order to get the total number of collisions over a period of time, instantaneous luminosity is integrated over time to get the total integrated luminosity L.

$$L = \int \mathcal{L} \tag{3.3}$$

By the end of Run 2, the LHC was able to deliver a total of 156 fb^{-1} of integrated luminosity. The ATLAS detector was only able to record a total of 147 fb^{-1} and qualified 140 fb^{-1} suitable for physics analysis [67]. The integrated luminosity recorded by ATLAS is shown in Figure 6.
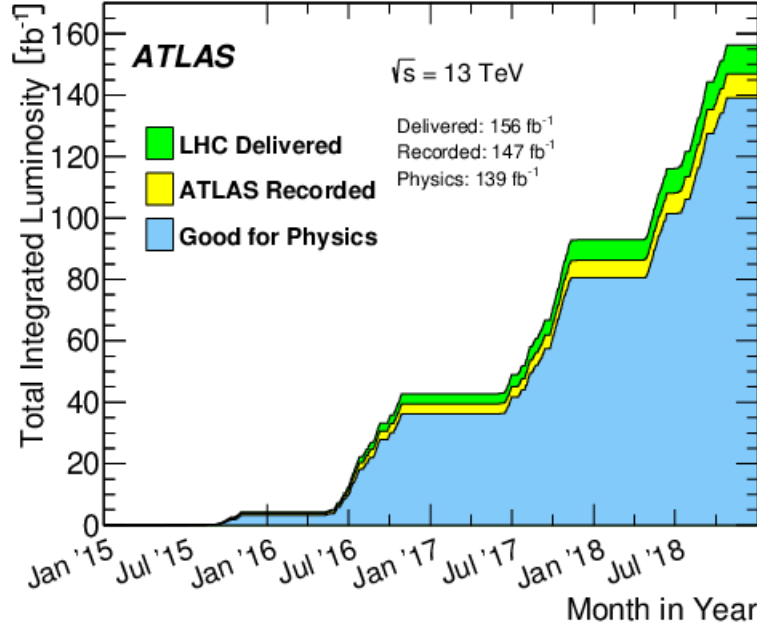
Figure 6: Cumulative integrated luminosity delivered and recorded by the ATLAS detector during Run 2 of the LHC between the years 2015 to 2018 with a CME of $\sqrt{s} = 13$ TeV

Dense proton bunches collide in the LHC beam line interaction points, meaning there are multiple pp interactions at one time. Not all interactions are considered and therefore are not saved. Typically, the most energetic collision per bunch crossing, called the hard-scatter, is analyzed. The symbol denoting the average number of pp collisions per bunch crossing is μ and is termed pile-up. The average pile-up for Run 2 is calculated to be $\langle\mu\rangle = 33.7$. The average pile-up over each data taken year of Run 2, along with the total average, can be seen in Figure 7 [67].

3.2   The ATLAS Detector

The ATLAS detector is a general purpose detector located 100 meter below CERN at interaction point 1 on the LHC ring. The purpose of this detector is to probe the SM using pp collisions and *Nucleus-Nucleus* collisions. Many types of physics analyses have been conducted using the ATLAS detector, such as testing predictions for BSM within the SM as well as precision measurements for its physical parameters such as particle mass and lifetimes. The high energies provided by the LHC allows physicists to probe very heavy particles such as the top quark and the Higgs boson. ATLAS
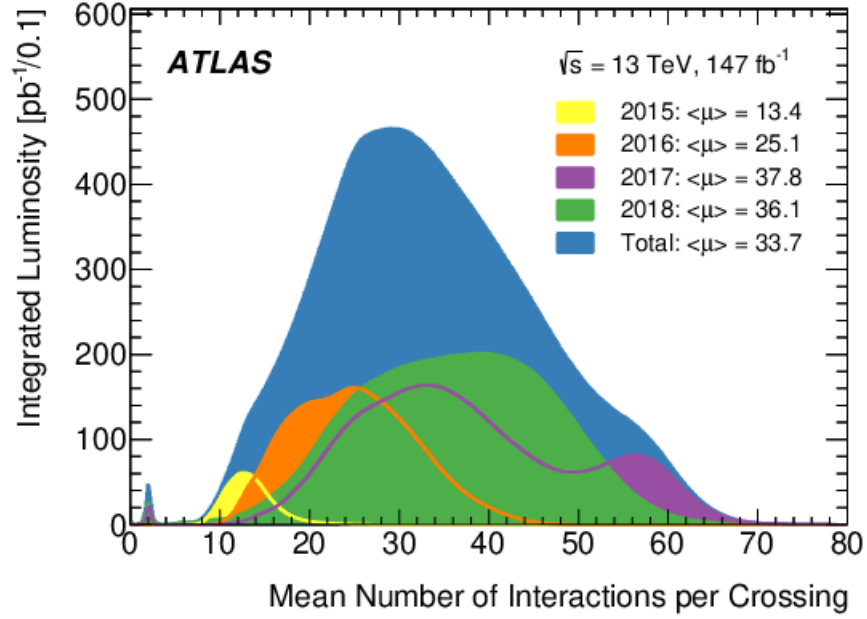
Figure 7: The μ distribution measured by the ATLAS detector during Run 2 [67]

is the largest detector located at CERN. The detector extends 25 meters in height, 44 meters in length and weighs 7000 tons. Figure 8 shows an overview of the detector.

The orange center of Figure 8 is the first of the sub-detectors called the Inner Detector (ID). Approximately 1000 particles are released every 25 ns, therefore high granularity is essential to differentiate the tracks of these particles. The ID is comprised of several parts, pixel and silicon microstrip (SCT) trackers, these are used in conjunction with the Transition Radiation Tracker (TRT.). The ID's main goal is to obtain the trajectories and paths of the particles that travel through it. Outside of the ID is a 2T solenoid magnet that bends the trajectories of the particles, which allows the momentum to be calculated. After the 2T magnet are the calorimeters which function as energy deposition detectors. There are two types of calorimeters used, an electromagnetic calorimeter (ECAL) in which electrons deposit their energy, the second is called the hadronic calorimeter (HCAL) where hadrons deposit their energy. Outside of these two calorimeters is the muon spectrometer (MS) which is needed for precise muon momentum measurements. All of these systems form a large cylindrical shape around the LHC beam pipe. End-caps are placed on the left and right side of the barrel which consist of wheel-shaped detectors.
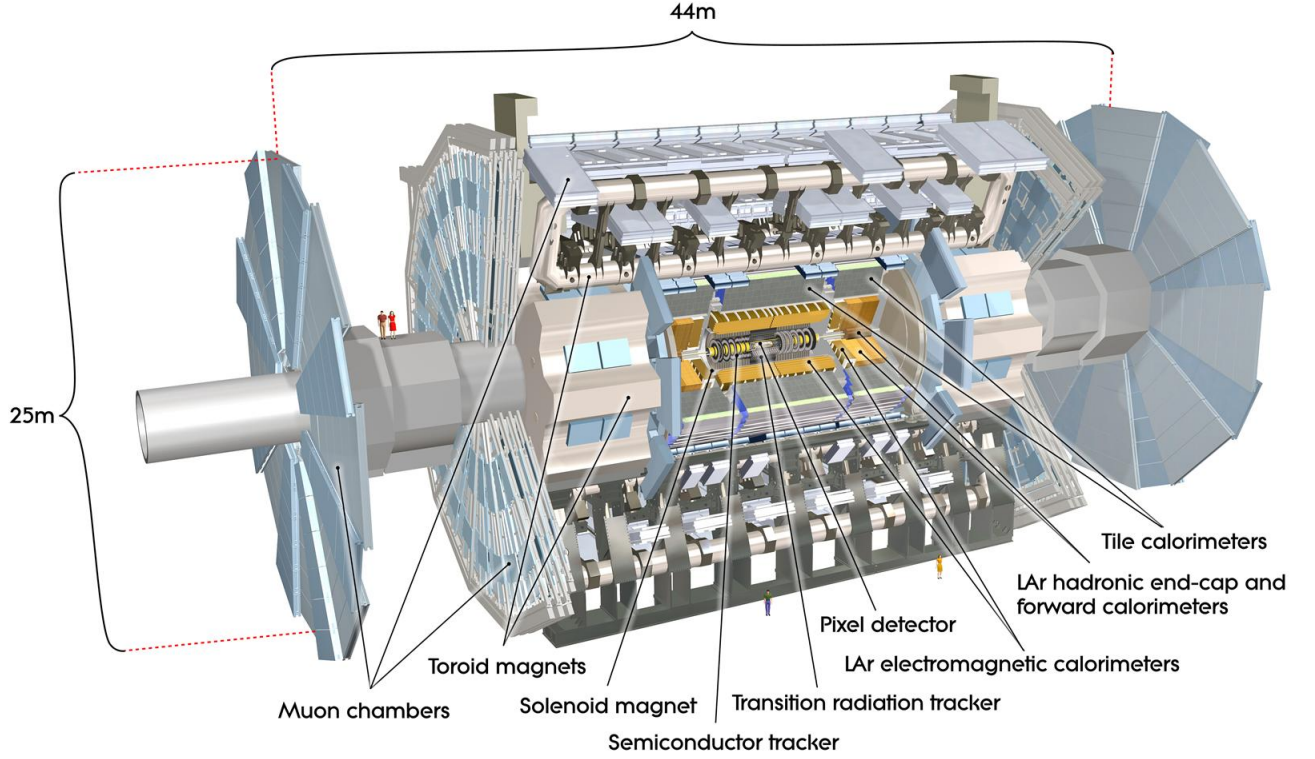
Figure 8: The ATLAS detector and its subsystems as of Run 2 [29]

3.2.1 The Coordinate System of ATLAS

A cylindrical and a right-handed Cartesian coordinate system are used in junction within the detector with the origin placed at the collision point. The x-axis points towards the center of the LHC, the y-axis points straight up and is used for the basis of the cylindrical coordinate system. The z-axis points directly down the beam-pipe. Two angles are used to define positions in the detector, the azimuthal angle ϕ and pseudorapidity defined as η. Pseudorapidity is defined in Eq. 3.4:

$$\eta = -\ln(\tan(\frac{\theta}{2})) \tag{3.4}$$

Where θ is the longitudinal angle measured between the z-axis and the trajectory of the particle. Pseudorapidity is an approximation of the variable rapidity and is commonly used due to its ease of calculation from the Cartesian coordinate system and is equivalent to rapidity for massless particles.

Rapidity, y, is defined as:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{3.5}$$

Rapidity, Δy, is invariant under Lorentz boosts. A benefit of using these variables is that η and ϕ provide a Lorentz invariant coordinate system with a distance measurement ΔR, defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.6}$$

Due to the physical design of ATLAS, it is impossible to cover the entire η range where highly boosted particles are able to leave through the beam-pipe, traveling very close to the z-axis. Table 3 shows the total pseudorapidity, η, range of all the sub-detectors.

sub-detector	η coverage
Inner Detector	± 2.5
Electromagnetic Calorimeter	± 3.2
Hadronic Calorimeter Barrel and end-cap	± 3.2
Forward Calorimeter	$3.1 < |\eta| < 4.9$
Muon Spectrometer	± 2.7

Table 3: Pseudorapidity of ATLAS sub-detectors [29].

Protons from the collision point only have momentum in the *z*-direction. Therefore the sum of the momentum in the transverse plane $\vec{P}_T = (P_x, P_y)^t$ of particles from the collision point must yield zero. Some particles, such as neutrinos or BSM particles, can escape the detector without depositing energy. The transverse momentum sum of all escaping particles can be determined from knowing that the total transverse momentum sum of all escaping particles must be zero. If a particle leaves without depositing energy, then the sum of the total transverse momentum will be negative and can be shown in the form $\vec{P}_T^{\text{miss}} = -\sum_{\text{detected}}^i \vec{P}_T^i$. The absolute value of $\vec{P}_T^{\text{miss}}$ is denoted as $E_{T\text{miss}} = |\vec{P}_T^{\text{miss}}|$. This value is referred to as missing transverse energy (MET).

3.2.2 The Inner Detector

The ID consists of three independent sub-detectors that are contained within a cylindrical sleeve with measurements of $\pm$3512 mm and a radius of 1150 mm within a solenoid magnet of 2T. The ID has a pseudorapidity range of $|\eta| \leq 2.5$ and a p_T threshold of 0.5 GeV [29]. The *pp* collision points are contained within the ID. The collision point that contains the highest p_T is considered the primary vertex. As the decay products travel from the primary vertex, after some time, decay themselves, creating a secondary vertex point. The ID was uniquely designed to provide robust pattern recognition and momentum resolution for both primary and secondary vertex measurements for charged particle's tracks. The three sub-detectors provide complimentary data acquisition. A decay product is first met by the pixel detector, containing discrete space-points from silicon pixel layers. It is then met with stereo pairs of silicon micorstrip (SCT) layers. Just beyond this is the TRT, containing many layers of gaseous straw tubes interwoven with transition radiation material. Figure 9 shows a layout of the ID, mapping its pseudorapidity coverage and the locations of each sub-detector.

The *pp* collisions impose a high-radiation environment which require stringent requirements for material composition and sensors of the ID sub-detectors. With the designed life-time of 10 years of operation during Run 2 at the LHC, the pixel inner vertexing layer must be replaced every three years. In order to subvert radiation damage over time, the silicon sensors are kept at low temperatures, about -5°C to -10°C resulting in coolant temperatures of -25°C [29]. The TRT is kept at room temperature. Figure 10 shows a drawing of each sub-detector and their sensors along with a transversed track of a charged particle.

The Pixel Detector

The pixel detector consists of three concentric layers of sensors in the ID barrel and three end-cap discs with sensors parallel to the transverse plane. The barrel layers are placed at a radii of 50.5 mm, 88.5 mm, and 122.5 mm as shown in Figure 10. The end-cap discs are located $|z| = 495$ mm-650 mm and cover r = 88.8 mm-149.6 mm giving an eta coverage of $|\eta| < 2.5$. Silicon sensors are
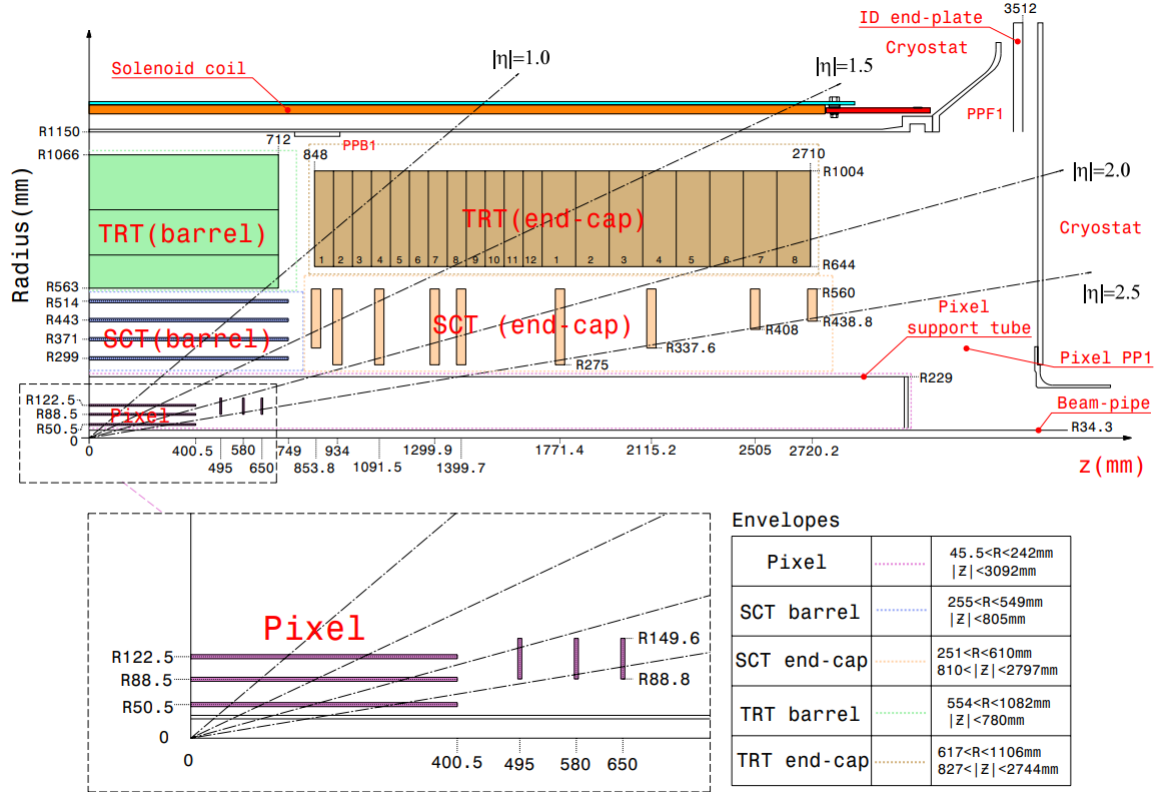
Figure 9: A quarter-section of the ATLAS ID showing its major elements with their active dimensions. [29]

pixelated with a pixel pitch of 50×400 μm^2 (90%) with the remaining being 50 μm × 600 μm^2. Equaling 47232 pixels per sensor. The short side of the pixel sensors are orientated along $r - \phi$ in both the barrels and the end-caps. The pixel orientation are 10 μm along the $r - \phi$ and 115 μm in the z (r) in the barrel (end-caps).

The SCT Detector

As shown in Figure 10, the SCT uses silicon strips distributed along four layers in the ID barrel and nine discs in each end-cap [29]. These layers cover ranges from $r = 299$ mm to $r = 514$ and cover $z < 749$ mm. The end-cap discs are placed at $|z| = 853.8$ mm and $|z| = 2720.2$ mm allowing up to $|\eta| \leq 2.5$ as seen in Figure 11. The strip modules are rectangularly shaped with 12 cm strips aligned in the z-direction. Trapezoidal strips are implemented in the barrel end-caps. The SCT strip pitch is 80 μm. Three-dimensional coordinates are found by setting two strips back-to-back
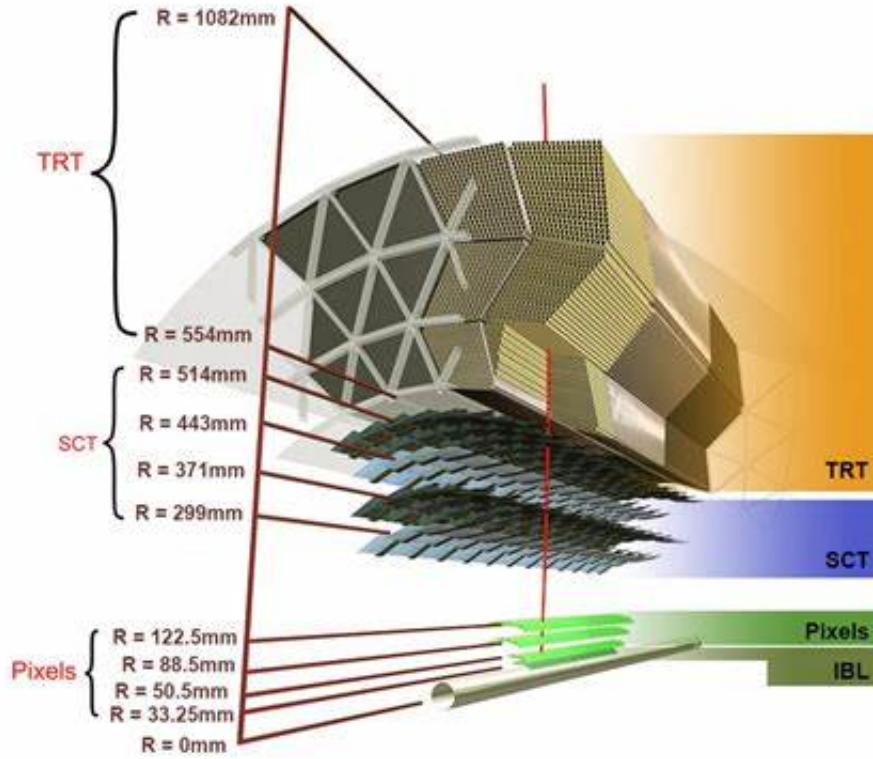
Figure 10: Drawing of the ID sub-detectors and their sensors. Showing the spacing of each of the sensor layers along with a transversed charged track at η=0.3. [29]

with an offset of 40 mrad. The intrinsic position of the SCT is 17 μm in the $r - \phi$ and 580 μm in z (r) in the barrel (end-caps)

The TRT Detector

The TRT is composed of 4 mm diameter polyimide drift straws containing a gas mixture of 70% Xe, 27% CO_2 and 3% O_2 set to 5−10 mbar over-pressure. The straws are made of two 35 μm thick multi-layer films bonded back-to-back. The bare 25 μm thick polyimide film is coated on one side with a 0.2 μm Al layer which is protected by a 5-6 μm thick graphite-polyimide layer [29]. The straws are stabilized using carbon fibers and cut to length of 144 cm for the barrel and 37 cm for the end-caps. For the these drift tubes, 31 μm diameter tungsten anode wires plated with 0.5−0.7 μm gold are used and are supported by the straw ends and the end plug. These are connected to the electronics and kept at ground potential. Cathodes operate at −1530V to give a gain of 2.5×10^4.
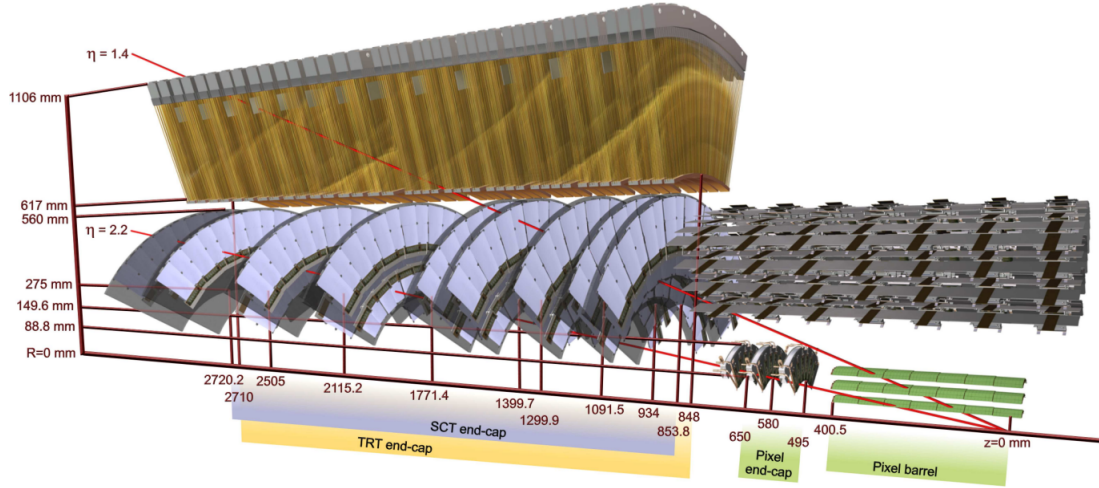
Figure 11: Drawing of the ID sensors of all three sub-detectors and their structural elements focusing on its end-caps transversed by two charged tracks. [29]

The gas within the straws are ionized as a charged particle interacts with it, releasing drift electrons. These electrons are collected at a time of 48 ns. The Xe-based gas mixture requires a re-circulating gas system with continuous monitoring of its quality. The TRT offers a particle position resolution of 130 μm which is much less than the previous two sub-detectors. The TRT also aids in the identification of charge particles by exploiting transition radiation.

3.2.3 The Calorimeter System

Surrounding the ID solenoid are both of the calorimeters that are responsible for measuring charged particles Energy. Each calorimeter consists of a number of sampling detectors that cover a full ψ-symmetry and coverage around the beam axis. The first calorimeter is the ECAL which detect depositions of electromagnetic particles such as electrons and photons. Outside of the ECAL sits the hadronic calorimeter, its responsible for measuring energy of any hadronic products from the collision. Both calorimeter systems use liquid argon as the medium for detecting. These calorimeters have alternate between two types of layers, an active layer and a sampling layer. The active layer interacts with charged particles, releasing secondary particles that the sampling layer then measures. A calibration is done and converts the deposited energy within the sampling layer to the

parent-collision particle's energy.

Electromagnetic Calorimeters
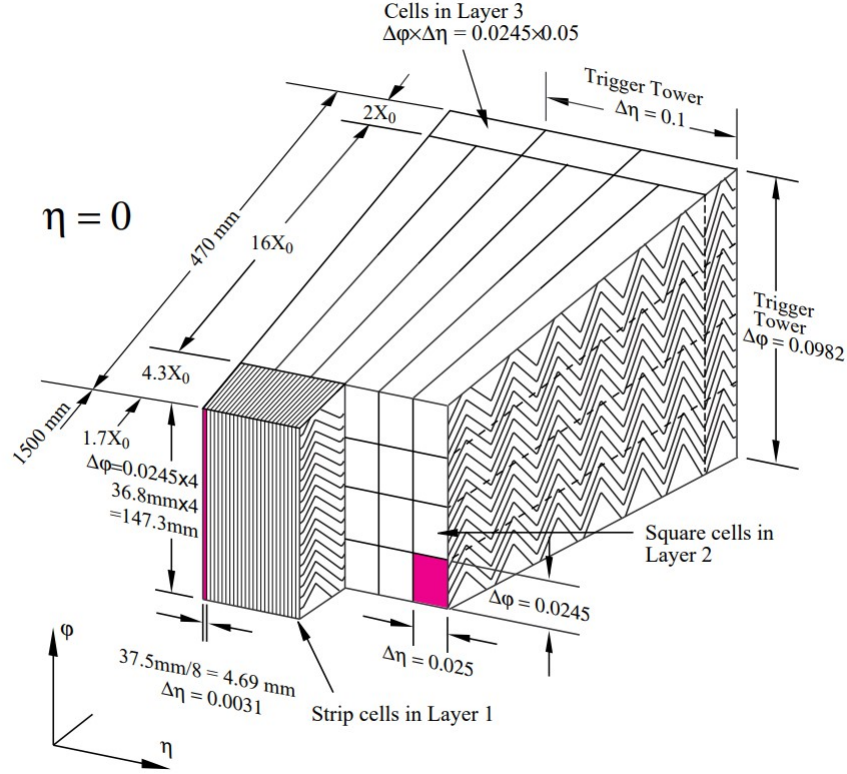
Figure 12: Sketch of a barrel module at η=0. The folds of the electrodes and each layer's granularity is shown [29]

The ECAL is just outside the solenoid of the ID and has a barrel shape with two end-caps. Inside this detector are plates of lead laid inside two sheets of stainless steel that act as an absorber, while liquid-argon acts as the sampling material. Three conductive copper layers are located in the gaps between the absorbers and act as electrodes. These electrodes and absorbers inside this calorimeter are folded in an accordion style. The folding radius varies in length to keep the liquid-argon gap constant. Both, the electrodes and absorbers, vary in radius with respect to the liquid-argon gap with the end-caps due to being parallel to the radial direction. There are two barrels that cover positive and negative η and are centered at the *z* axis, called the LAr electromagnetic barrel. One half of the barrel covers $0 < \eta < 1.475$ while the second half covers $-1.475 < \eta < 0$. The length

of each barrel is 3.2 m, the inner and outer diameter are 2.8 m and 4 m respectively, weighing a total of 57 tons. Figure 12 shows a sketch of a barrel module at $\eta = 0$ showing each of the three layers and their granularity.

The end-caps of the EM calorimeters (EMEC) consist of two wheels, one on each side of the barrels. Each wheel is 63 cm thick and weighs 27 tons. These wheels have an η coverage of $1.375 < |\eta| < 3.2$. A liquid-argon sampler is placed in the region between the wheels of the EMEC and the ECAL to improve the energy measurement within the region which has an η coverage of $1.5 < |\eta| < 1.8$. Each EMEC consists of two co-axial wheels located in between the inner and outer wheel. These are 3 mm wide and located at $|\eta| = 2.5$.
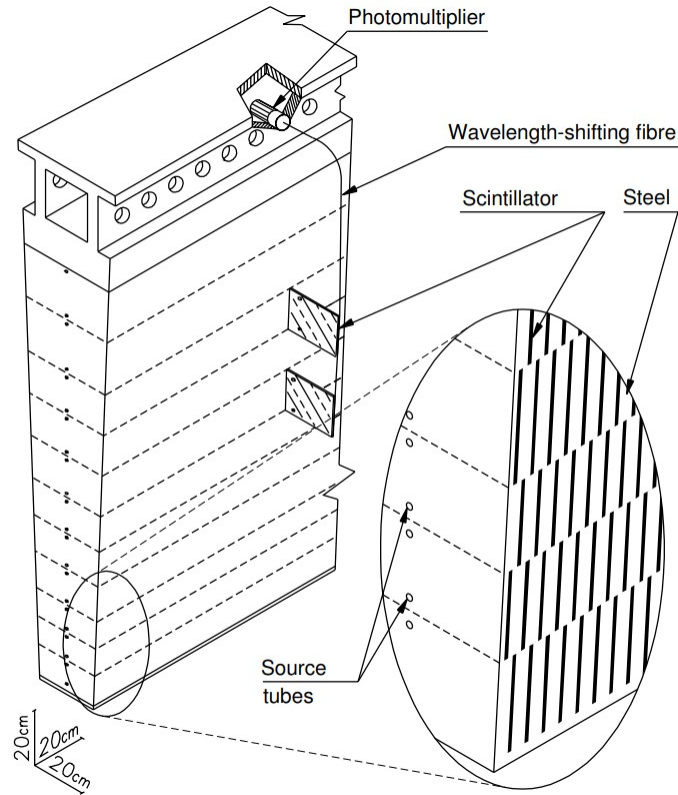
Figure 13: Sketch of the tile calorimeter showing the alternating steel plates and scintillator material attached to PMTs. [29]

Hadronic Calorimeters

There are three sub-detectors within the HCAL system, the tile calorimeter, the liquid-argon hadronic end-cap calorimeter (HEC), and the liquid-argon forward calorimeter (FCal). The tile calorimeter is a sampling calorimeter that uses steel as the absorber and an active medium as a scintillator. It's located just outside the ECAL and covers an η region of $|\eta| < 1.7$ and is divided into three parts, the center barrel extends 5.8 m in length while the two outer barrels extend 2.8 m. The orientation of the scintillator tiles are radial and normal to the beam line. Attached are wavelength-shifting fibers that extend into readout photomultiplier tubes (PMT). A sketch of the tile calorimeter can be seen in Figure 13. The output from the PMTs are what's recorded.

The HEC schematic is shown in Figure 14. This module is an end-cap calorimeter consisting of copper and liquid-argon sampling with a flat plate design, cover a range of $1.5 < |\eta| < 3.2$. These end-caps share the cryostats with the EMEC and the FCal. It consists of two wheels in each end-cap, front wheel (HEC1) and the back wheel (HEC2). Each wheel contains two longitudinal sections constructed of 32 identical wedge-shaped modules. The FCal provide an η coverage of $3.2 < |\eta| < 4.9$. These modules are located at high η, approximately 4.7 m from the collision point, exposing them to high particle fluxes. This fact resulted in much smaller liquid-argon gaps to avoid ion build-up and faster signal readout. Figure 15 shows an illustration of the entire HCAL section in ATLAS.
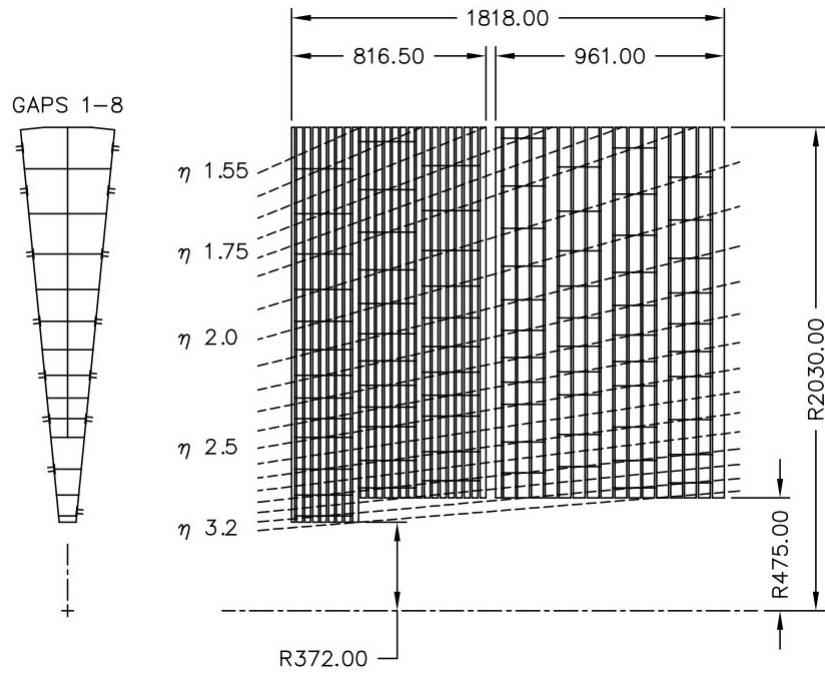
Figure 14: Hadronic calorimeter end-cap. Electrode readouts indicated by dashed lines . [29]
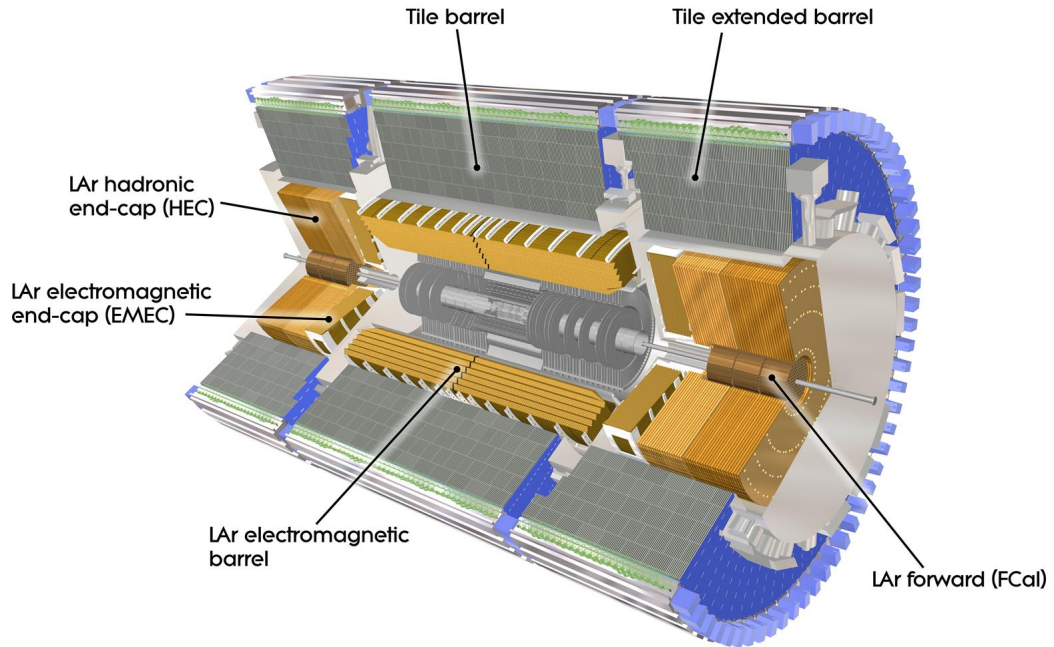
Figure 15: Illustration of the entire HCAL system in ATLAS. [29]

3.2.4   The Muon System

The final system is the muon spectrometer which is designed to detect charged particles exiting the calorimeters and to measure their momentum. Muons lose minimal energy when traversing through matter and therefore are depositing minimal energy in the ECAL. The muon system triggers on particles with the pseudorapidity of $|\eta| < 2.4$ and can track up to $|\eta| < 2.7$. The system consists of three concentric cylindrical shells of gas-filled detectors around the beam axis with radii of 5 m, 7.5 m, and 10 m. The two end-caps of the muon chambers form large wheels, perpendicular to the z-axis and are located at $|z| \approx 7.4$ m, 10.8 m, 14 m and 21.5 m from the collision point [29]. Figure 16 shows the overall arrangements in two cross-sections of the muon system. The muon trajectories are bent using large superconducting toroidal magnets with a generated field mostly perpendicular to the traversing muons.

The muon gas chambers are filled with 93% Argon gas and 7% carbon dioxide, inside the chambers are Monitored Drift Tubes (MDT) that perform the precision momentum measurements. The layers consist of three to eight layers of MDTs and can achieve an average drift resolution 80 μm per tube. In the forward region, $2 < |\eta| < 2.7$, Cathode-Strip Chambers (CSC) are installed within the innermost tracking layer. These are used for their high rate capability and time resolution due to the increase of particle flux. The CSC have cathodes perpendicular to the MDTs and the charge induced on the cathode is measured for the signal. The resolution of each chamber is 40 μm in the bending plane (due to the toroidal magnet) and 5 mm in the transverse plane. The muon system is completed with a high precision optical alignment system and is paired with track-based alignment algorithms.
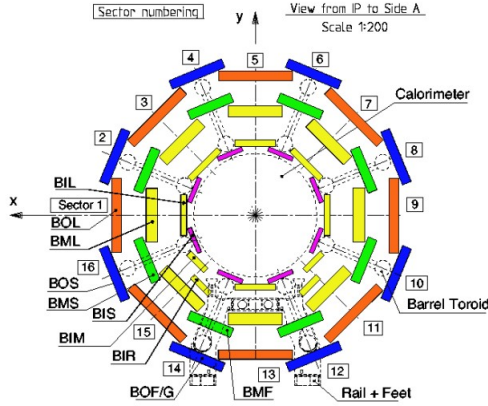
(a) Cross-section of muon system
perpendicular to the beam-axis
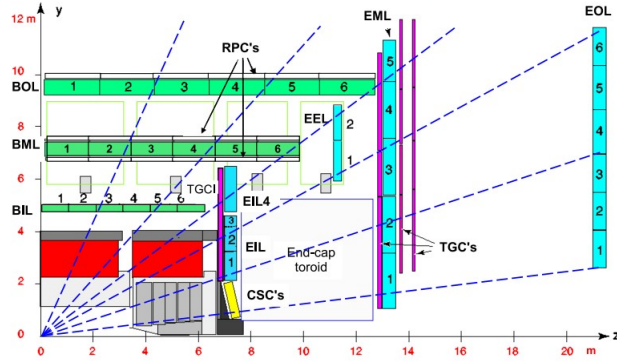
(b) Cross-section of muon system in a plane with the
beam-axis

Figure 16: Cross-sections of the muon system in the ATLAS detector. Figure (a) shows the three concentric cylindrical layers of 8 large and 8 small chambers. Outer diameter being 20 m. Figure (b) shows a planar view of the muon system, non-bending muon tracks are shown with dashed lines [29]

3.2.5   Trigger and Data Acquisition

pp collision rate inside ATLAS occurs at a very high frequency. Bunch crossings occur every 25 ns which equates to a rate of approximately 40 Mhz [29]. The amount of data produced in these *pp* collisions exceeds the storage capabilities, additionally, most bunch crossings feature only low *pp* momentum transfers and are therefore not interesting to analyze, let alone store. Therefore a complex selection system is implemented within ATLAS called the Trigger and Data Acquisition (TDAC) system [12]. This system is necessary to only store events that are of interest, reducing the necessity for storage facilities.

The trigger system consists of three levels of trigger systems. They are Level-1 (L1), Level-2 (L2), and event filter. The L2 and event filter triggers compose the High-Level Trigger (HLT). The L1 trigger is implemented using custom-made hardware, while the HLT is based on the latest software algorithms. The highest acceptance rate of the L1 trigger is 75 kHz and the decision making must reach the front-end electronics within 2.5 µs. This trigger mainly searches for signatures from high-p_T muons, electrons, photons, jets, τ-leptons that decay into hadrons, and MET using information collected by the calorimeters and muon spectrometer. Figure 17 shows a diagram for the L1

trigger. This information is passed to the L2 trigger which reduces the event rate to below 3.5 kHz. The L2 separates this information into Regions-of-Interest (RoI) based on coordinates, energy and types of signatures. The event filter uses offline analysis procedures to further reduce the dataset by down to approximately 200 Hz by making decisions on fully built events. However, the average acceptance rate of the HLT is 1.2 kHz, corresponding to a data rate of 1.2 GBs^{-1} to be written to disc [12].
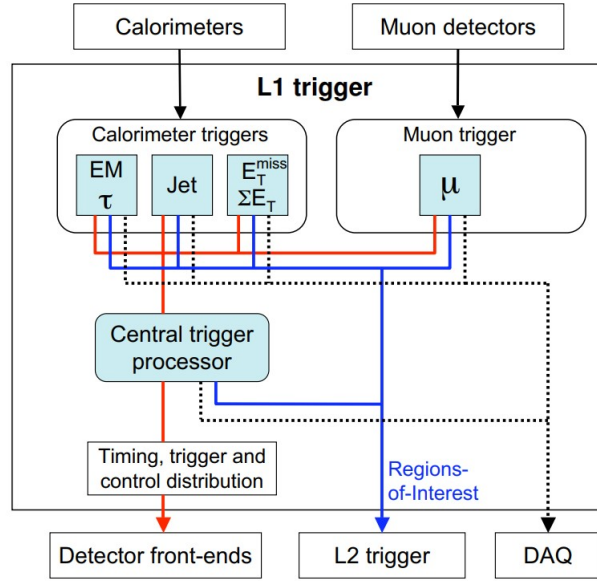
Figure 17: Block diagram of the L1 trigger in the ATLAS detector showing data paths to the front-end, L2 trigger and DAQs system. [29]

3.3 The High-Luminosity LHC and the ATLAS Detector Upgrade

There are two crucial factors in defining rare physics searching capabilities, one is the CME and the other is integrated luminosity. The CME dictates the production cross-section for each physics process, allowing for the produced particle to shift from off-shell (produced at a non-theorized mass) to on-shell (produced at theorized mass). Increasing the luminosity equates to increasing the total data production, therefore increasing the amount of rare events one may be looking for. The LHC was initially designed to run at a CME of 14 TeV in which it has obtained a value of 13.6 TeV during Run 3. The current luminosity obtained as been 140 fb^{-1} during Run 2 and an estimated 300

fb^{-1} during the course of the ongoing Run 3.

The HL-LHC is aimed to fully utilize the capabilities of the LHC with a running lifetime of 12 years (called Run 4). It is designed to increase the number of collisions per bunch crossing from $\langle\mu\rangle = 33.7$ to $\langle\mu\rangle = 200$. The expected peak luminosity achieved by the HL-LHC is 5×10^{34} cm^{-2}s^{-1} The expected integrated luminosity of each year is 250 fb^{-1}, with a total of 3000 fb^{-1} after the 12 year run time [18]. This massive upgrade to the LHC requires systems to be upgraded and some to be fully exchanged since these systems are vulnerable to breaking down due to high radiation exposure. This increase in radiation is also taken into account within the ATLAS detector, realizing the current infrastructure within the detector is unable to withstand this amount of radiation for the running time of Run 4. This high information dense environment now imposes a few new challenges, such as maintaining high granularity data readout and the ability for the electronics to maintain this operative power under such high radiation doses. In order to obtain this new feat, the ID will be completely replaced with a new detector sub-system called the Inner Tracker (ITk) as a part of the ATLAS Upgrade. Part of the work in this thesis focuses on preliminary studies of the latest particle identification algorithms implementing the new geometry imposed by the ITk through computer simulations.

3.3.1   The Inner Tracker

The new tracking detector is designed for a 10 year life span of operation at instantaneous luminosity of 7.5×10^{34} cm^{-2}s^{-1}, 25 ns per bunch crossing and a total integrated luminosity of 3000 fb^{-1} over the entire lifetime [30]. The current solenoid magnet being used will remain in place and provide a magnetic field of 2T as per the previous runs. One of the major improvements (besides being radiation hard) is that the new design will allow data collection of quality events of up to $|\eta| < 4.0$. This is achieved through a complex system of silicon barrel layers and disks or rings along with inclined pixel modules to have better coverage between the barrel and the end-caps of the ITk. The barrel will extend $\eta \approx \pm 1$ in which the current ID barrel only has a $\eta \approx \pm 0.6$ coverage.

ITk Layout

In the central area in the ITk barrel, sensors are arranged in cylinders around the beam axis. The first five layers are pixel modules followed by two short-strip layers of stereo modules then followed by two even longer strip stereo modules. The forward η regions will be covered using six strip disks on either side and several pixel rings.
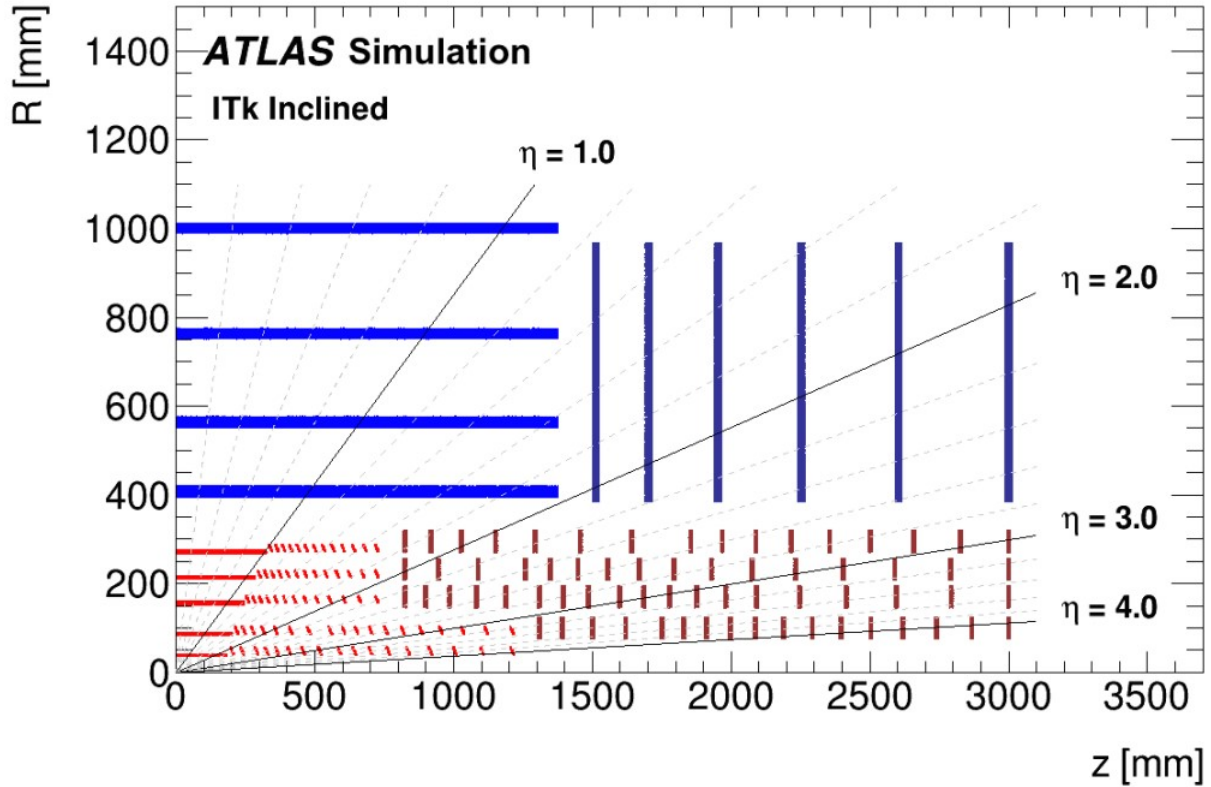
Figure 18: The ITk layout. The pixel detectors are shown in red and the strip Detector in blue. [30]

The strip detectors seen in blue in Figure 18 consists of four-layer barrel sections and one end-cap on each side containing six disks. This strip system covers η ≈ ±2.7 in the barrel region and sits ± 1400 mm along the *z*-axis. The inner layer strips extend a length of 24.1 mm while the outer layer strips are 48.2 mm. All strips within the ITk have a pitch of 75.5 μm. Table 4 shows layout parameters of the strip detectors within the detector barrel.

The strips in the six end-cap wheels on each side are radially distributed, pointing towards the *z*-axis. The strips vary in length within these end-caps to optimize total strip occupancy, starting

ITk Barrel Strip Detector Layout Parameters					
Layer	Radius [mm]	Channels in ϕ	Strip Pitch [mm]	Strip Length [mm]	Tilt Angle [∘]
0	405	28×1280	75.5	24.1	11.5
1	562	40×1280	75.5	24.1	11.5
2	762	56×1280	75.5	48.2	10
3	1000	72×1280	75.5	48.2	10

Table 4: Layout Parameters of the ITk Strip Detector barrel. Each strip is 2.8 m long. [30].

at 19.0 mm closer to the beam axis, varying up to 60.1 mm in the outermost regions. The exact locations of the end-cap disks and strips are shown in table 5.

Aside from the strip detectors within the ITk, there are pixel detectors as shown in red in Figure 18. The pixel detector consists of five barrel layers and four end-cap ring layers. The layout provides a full eta coverage of $|\eta| = 4$. The design of the pixel detector allows it to be operating the full lifetime of 12 years. The barrel layer of pixel sensors are placed tangentially of the constant radius of the cylindrical barrel shape. The sensors in the forward parts of the barrel are inclined at an angle of 56°. Currently, each pixel size is nominally set to 50×50 μm^2 with a thickness of 100 μm for the two most inner barrel layers and 150 μm for the outer two for simulation.

3.3.2 B-Tagging and Vertex Reconstruction for the ITk

Event reconstruction is crucial to the integrity of all physics conducted at the LHC. In order to effectively reconstruct events, the energy deposition and associated tracks must be recorded for the specified event to be rebuilt. As stated earlier, at the HL-LHC it is expected to obtain a pile-up of $\langle \mu \rangle = 200$. This means the mean separation of primary vertices will be approximately less than 1 mm within the space of the ITk barrel. Meaning it's not possible for all primary vertices of each event to be reconstructed independently. Therefore, it's crucial that all high energetic events coming from a common vertex must be identified to a high efficiency. Vertex reconstruction within such an environment imposed strict requirements on tracking resolution close to the particle interaction points. The goal is to have reconstructed vertices for $t\bar{t}$ events to be greater than 0.95 and that the $t\bar{t}$ decay product is associated to the correct vertex at a rate greater than 0.90. The b-tagging versus light-jet rejection should be optimized for each layout and at minimum should match the

performance of the ID. There is a much more in depth discussion on particle identification and event simulation in chapters 4.1 and 4.7 respectively.

3.3.3 High Granularity Timing Detector

The High Granularity Timing Detector (HGTD) is a disc shaped detector that will be placed outside the ITk. It will be added in front to the end cap and forward calorimeters at $|z|$= 3.5m. It is composed of two forward and backwards disks with central half rings and stave concepts with a total area of 6 m^2 of silicon sensors. The HGTD offers a new and powerful technique to overcome the obstacle of pile-up. It takes advantage of the time spread of interactions that occur close in space but spread apart over time. This provides a time resolution with a precision measurement of up to 30 ps and will cover an 2.4 $<$ $|\eta|$ $<$ 4.0 where pile-up is rejected [36]. Figure 19 shows the assembly diagram of the HGTD
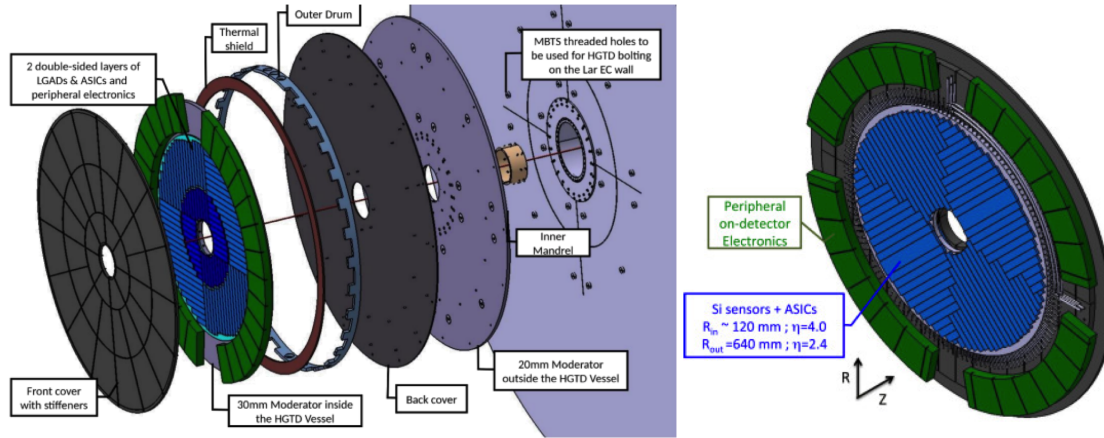
Figure 19: The assembly of the HGTD detector[36]

The sensors that are built within the HGTD are referred to as Low Gain Avalanche Detectors (LGADs). These sensors will have an active thickness of 50 μm and the pixel area covers 1.3 $\times$ 1.3 mm^2. The sensors will be integrated to an electronic circuit that is currently being developed to meet the requirements of the necessary timing resolution and radiation hardness [36]. The LGAD sensors are n-on-p silicon detectors with an internal gain. To obtain this gain, a highly doped, p-layer is added just below the p-n junction. This doped region creates a very high electric field and

will induce an avalanche of the electrons and thus creating additional electron-hole pairs. Figure 20 shows an illustration of the HGTD. The blue represents the active LGAD region and the green is an inactive off-detector region.
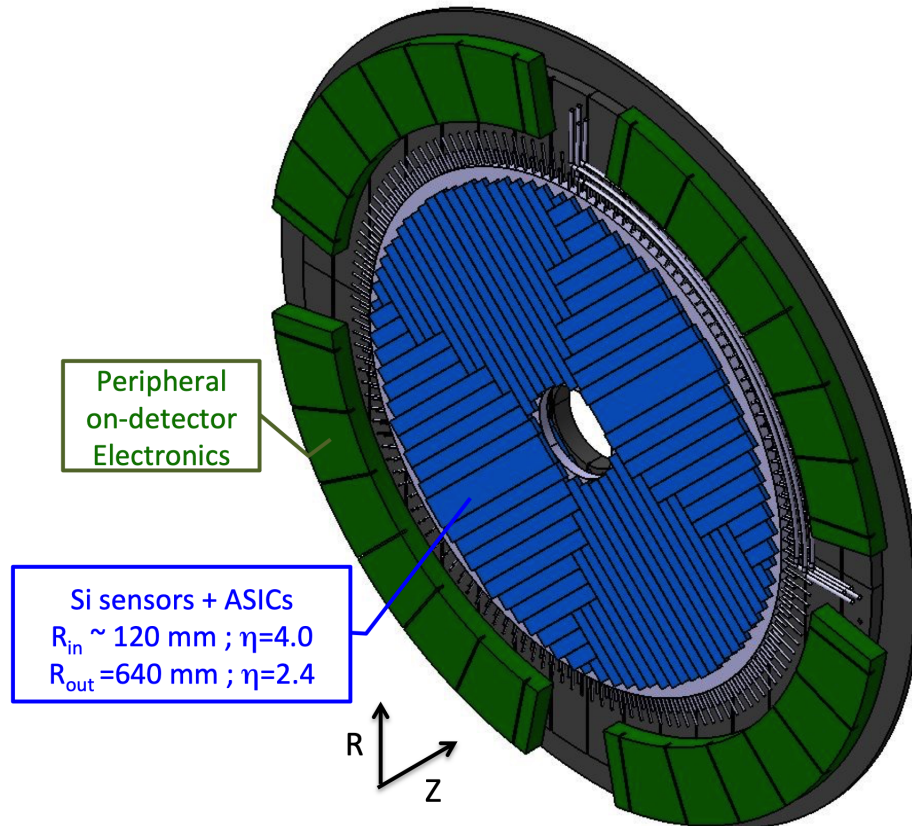
Figure 20: An illustration of the HGTD showing the LGAD sensors (blue) and an inactive region off-detector (green) [36]

ITk Barrel Strip Detector Layout Parameters			
Ring/Row	Inner Radius [mm]	Strip Pitch [mm]	Strip Length [mm]
Ring 0 Row 0	384.5	75.0	19
Ring 0 Row 1	403.5	79.2	24
Ring 0 Row 2	427.5	74.9	29
Ring 0 Row 3	456.4	80.2	32
Ring 1 Row 0	489.8	69.9	18.1
Ring 1 Row 1	507.9	72.9	27.1
Ring 1 Row 2	535	75.6	24.1
Ring 1 Row 3	559.1	78.6	15.1
Ring 2 Row 0	575.6	75.7	30.8
Ring 2 Row 1	606.4	79.8	30.8
Ring 3 Row 0	638.6	71.1	32.2
Ring 3 Row 1	670.8	74.3	26.2
Ring 3 Row 2	697.1	77.5	26.2
Ring 3 Row 3	723.3	80.7	32.2
Ring 4 Row 0	756.9	75.0	54.6
Ring 4 Row 1	811.5	80.3	54.6
Ring 5 Row 0	867.5	76.2	40.2
Ring 5 Row 1	907.6	80.5	60.2

Table 5: Main layout parameters for the strip detector end-caps. [30].

CHAPTER IV

OBJECT RECONSTRUCTION AND EVENT SIMULATION IN ATLAS

The previous chapter went into detail on how the ATLAS detector works through its various sub-detectors and and systems. Collisions occur at the beam axis, creating a shower of particles into the detector, depositing energy on an object's corresponding detector and having its tracks mapped by several systems within the ID. These recorded signals from *triggered* events are used to reconstruct the event using complex algorithms.

Figure 18 shows a slice of the ATLAS sub-detectors with several different objects from a single event depositing energy in corresponding sensors. The trajectories of charged particles (tracks) are reconstructed using the ID [29]. Muons are reconstructed using associated tracks measured within the muon spectrometer and tracks left within the ID [61]. Electrons are reconstructed using tracks detected within the ID along with energy deposits within the ECAL [31]. Photons will also deposit their energy within the ECAL, but do not leave any tracks, thus being able to separate both electromagnetic objects.

From these collision events are objects known as jets, i.e., hadron cones due to hadronization of quarks and gluons as seen in Figure 21 as the red deposits in the blue HCAL. Due to hadronization, multiple tracks are associated with jets which is why jets are reconstructed and not individual hadrons. Jets can be caused by several different hadrons and so we separate these types into, what we call, flavors. Jets caused by *b*-hadrons are called *b*-jets, and likewise, *c*-hadrons cause *c*-jets. The lighter quarks (up, down and strange as discussed in Section 2.2) are difficult to differentiate due to their similar size in mass. We group these hadrons into a single category of jets called *light*-jets. As jets propagate outward they create a cone like shape within the HCAL, covering an angular area. Jets that cover a large angular area are considered large-radius jets [13] which contain interesting

51

physics and are a focused object for the preliminary analysis in the last chapter of this thesis.
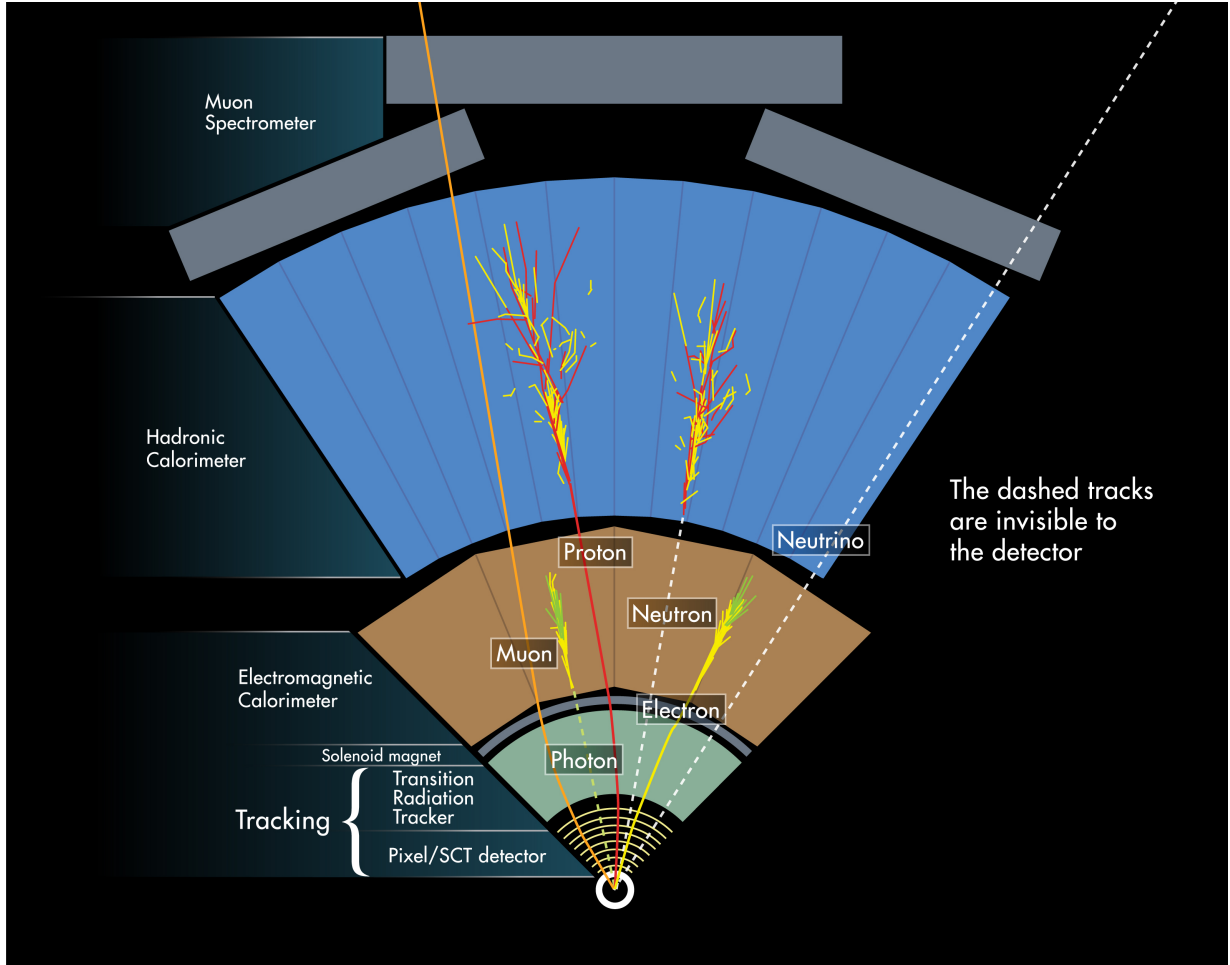
Figure 21: Planar slice of the ATLAS detector showing different energy deposits of several physics objects [83].

4.1   Track and Vertex Reconstruction

Almost all physics objects detected by ATLAS deposit small amounts of energy within the ID sensors, each *hit* is then used to reconstruct the object's track(s). Tracks are seeded by groups of three, one from the IBL, another from the pixel detector and lastly the SCT [9]. Within the pixel and SCT detectors, the track association begins with clustering raw measurements. A connected component analysis (CCA) [1] groups pixels and strip detectors together that share a common edge or corner if energy has deposited on them above a certain threshold. These combined clusters

52

are then referred to as *space-points*. The average size of a pixel cluster is about 2 pixels in the r-ϕ plane and 1 to 3 pixels in the longitudinal direction with increasing η. Clusters from single charged particles are *single-particle* clusters, clusters from multiple charged particles are referred to as *merged* clusters and clusters from multiple tracks that are not compatible are called *shared* clusters. A Kalman filter [60] is applied to these *merged* clusters to maximize track purity for objects of interest. Particles often share a hit, in this case a quality score is assigned to each track candidate. The score is determined by the number of ID layers with a hit, a track fit χ^2 and the logarithm of the track momentum to suppress low p_T track objects [9, 58]. For clusters assigned to multiple tracks in the merged clusters category, an ambiguity solver is applied. For the track selection criteria, a track cannot be associated to no more than two shared clusters, cannot be over $p_T > 400$ MeV and $|\eta| < 2.5$. Using the hits in the TRT, the final track candidates are mapped out and a final selection occurs through track fitting.

Through the multiple *pp* collisions in a bunch crossing, there is typically a single collision that yields objects of interest, this is called the hard-scatter. The main vertex of the hard-scatter and the secondary vertices from its decay products are mapped through the tracking procedure by applying track criteria. Multiple vertices are found on the z-axis from luminous peaks, called seed vertices. Here is where the track algorithm and its criteria are applied to find the primary vertex of the hard-scatter.

4.2   Electrons

Electrons are reconstructed using the energy deposits within the ECAL and tracks within the ID. The reconstruction comes in several steps. The first is combining the energy clusters within the ECAL into EM-topo clusters. Tracks are then matched to these clusters, creating superclusters. Electrons are then defined from these superclusters [43].

Proto-clusters are first defined within the ECAL and HCAL using a set of noise thresholds to discriminate from background noise. The clusters found in the HCAL are not used in the electron reconstruction process, but are used to reduce background noise from pile-up. The cell clusters are

required to have a significance $|\zeta_{\text{cell}}^{\text{EM}}| \geq 4$ as defined in Eq. 4.1 [43]:

$$\zeta_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise,cell}}^{\text{EM}}} \tag{4.1}$$

$E_{\text{cell}}^{\text{EM}}$ is the cell energy deposition and $\sigma_{\text{noise,cell}}^{\text{EM}}$ is the expected cell noise. This noise is the known electronics noise and an estimate of the pile-up noise corresponding to the instantaneous luminosity of Run 2. The algorithm then iterates through these clusters, ignoring the first of the ECAL in order to suppress noise. These proto-clusters are then paired with neighboring clusters with a significance of $|\zeta_{\text{cell}}^{\text{EM}}| \geq 2$. These neighbors are then used as seed clusters of the next iteration to further find neighboring cells within the cluster. Once all the neighboring clusters with $|\zeta_{\text{cell}}^{\text{EM}}| \geq 2$ are found, they are then merged into a larger cluster. Proto-clusters with more than two or more local energy maxima, they are split into separate clusters. A maxima is considered when it has $E_{\text{cell}}^{\text{EM}} > 500$ MeV and at least four neighbors with smaller signal.

Tracks are then matched to the proto-clusters using information from the ID as described in Section 4.1 [4]. Electrons lose a significant portion of their energy within the ID compared to other charged particles. To ensure efficiency in electron reconstruction, a loose track association criteria is applied. The conditions are $|\eta_{\text{track}} - \eta_{\text{cluster}}| < 0.05$ and $-0.10 < q(\phi_{\text{track}} - \phi_{\text{cluster}}) < 0.05$ where q refers to the charge of the track. An ordering is applied to cases when multiple tracks are matched to a single cluster. Tracks found in the pixel detector are the most preferred, then tracks within the SCT with none in the pixel detector. Then the best ΔR between the tracks and the cluster chosen. The track with the best score is then matched to the ECAL cluster and is used in the further steps of electron reconstruction.

Superclusters are then defined from these chosen EM topo-clusters. The seeds chosen to form the superclusters are selected by requiring $E_{\text{T}} = \sqrt{m^2 + p_{\text{T}}^2} \geq 1$ GeV and matched track fulfilling quality criteria. These superclusters are then extended by adding topo-clusters within $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$, these are considered *satellite* clusters and are assumed to be secondary EM showers from the initial electron or photon. But for electrons, this distance is extended to $\Delta\eta \times \Delta\phi =$

0.125 × 0.3 and the seed and the adjacent topo-cluster must share a matched track. These steps rely heavily on tracking information to discriminate between radiative photons or low energy electrons from pile-up or noise.

Once the superclusters are built, an initial energy calibration and position correction are applied to them. Tracks are matched to these supercluster while conversion vertices are matched to photons since they do not leave tracks within the ID. There are events with clear cut identification of electrons and photons, i.e., superclusters with defined tracks (electrons) and superclusters with no defined tracks (photons). But there are cases when there is ambiguity between both objects, when this occurs a classification process is applied to determine the type of object. The initial suepercluster calibration occurs before the track matching and relies on energy calibrations of the electrons and photons to simulations [3]. This process is referred to as energy scale calibration. The energy resolution of electrons found in data are used to calibrate simulations of $Z \rightarrow ee$ events. These simulated events are then used to derive the energy scale and resolution calibration factors. The invariant mass of two electrons in data and within these simulated events after the energy resolution is applied can be found in Figure 22 with good agreement.
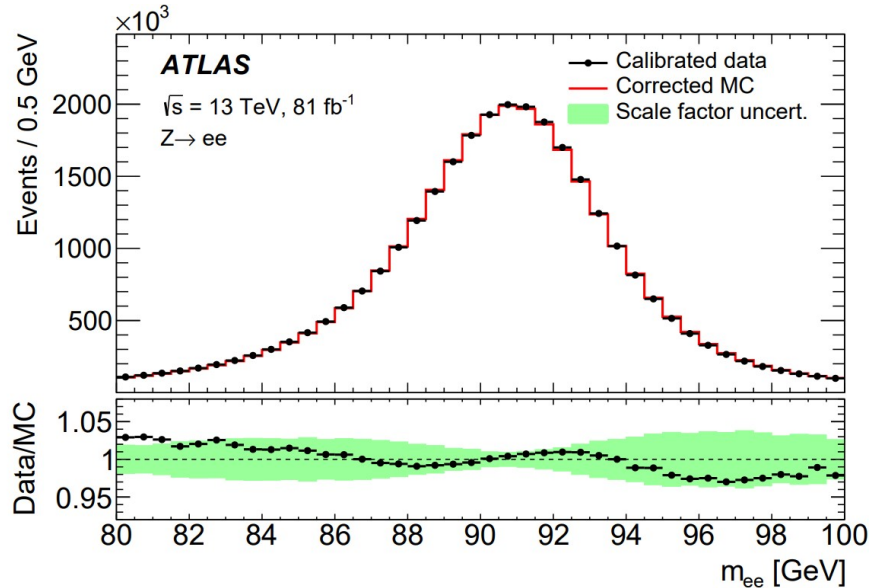
Figure 22: Comparisons of simulated events and data after calibration and resolution corrections applied for electrons [43].

4.3 Muons

Muons deposit little energy as it transverses the ATLAS detector, thus its reconstruction relies heavily on associated tracks left within the ID and the (MS) along with characteristic energy deposits in the calorimeters. The track reconstruction is for the MS is independent of the track reconstruction in the ID as described in section 4.1, though the reconstructed muon depends on both [61].

Tracks within the MS are identified as short straight-line local segments reconstructed from hits in individual stations. These segments are identified through a process known as a Hough transform [71]. These segments are combined into preliminary track candidates using a loose constraint of the impact parameter (IP, discussed in Section 4.5.1) and a parabolic trajectory that includes the first-order approximation of the track bending due to the magnetic field. These calculations are combined with precision information of a second coordinate from the trigger detectors, creating a three-dimensional track. This track then has a global fit χ^2 fit of the muon trajectory through the magnetic fields within the MS. The hits that are not contained within the fit of the muon trajectory are removed. Hits are added onto the fit trajectory that were not originally there, from these aligned hits, another χ^2 fit is applied. Less quality tracks are removed when they share a fraction of hits with a high quality track in order to resolve ambiguities. The final track is then re-fitted with a IP constraint that takes into account the energy loss in the calorimeters and an extrapolated trajectory back to the beam line [61].

This fitting procedure using hits within the MS and ID are applied in several different reconstruction strategies.

- Combined (CB): Combined muons are identified by matching MS to ID tracks and applying a combined fit as previously discussed. For muons with $|\eta| > 2.5$, MS track segments are combined with hits from the pixel detector and SCT detector. These boosted muons are considered a subset called silicon-associated forward (SiF) muons

- Inside-Out Combined (IO): Muons reconstructed using this method uses an algorithm which extrapolates the ID hits to the MS and searches for at minimum three loosely associated hits.

Between these loosely associated hits, the small energy loss within the calorimeters are used to verify. A Combined track fit is then applied.

- Muon-Spectrometer Extrapolated (ME): These are muon tracks within the MS but cannot be associated with any hits within the ID. If this is the case, the tracks are extrapolated all the way back to the beam line. Because of this, this extends the MS coverage to $|\eta| = 2.7$.

- Segment-Tagged (ST): ST requires at least one MS segment to be associated with an extrapolated ID track.

- Calorimeter-Tagged CT: These muons are identified by taking the extrapolated tracks in the ID through the calorimeters. Muons deposit minimum-ionizing energy in the calorimeters. If CT was used to identify the muon, the parameters are obtained directly through the track fit.

After reconstruction, high quality muon candidates are then selected using criteria utilizing the sub-detectors. The set of requirements on the identified muons given its type as stated above is referred to as its selection *working point* (WP). The definition of muon WPs depends on the type of analysis involving final state muons. The standard WPs designed to cover needs of most analyses vary in increasing purity and decreasing muon tagging efficiency. The WPs are *loose*, *medium*, and *tight*. Here, muons passing the *medium* WP is a subset of the *loose* WP. These were defined to help optimize different physics analyses. The *medium* WP is suitable for most general analyses where the *loose* was optimized for Higgs boson decays into a four muon final state [61].

Muon correction factors are obtained through simulation events of $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$. The momenta in the simulation are corrected to those obtained through data. Multiplicative correction factors are obtained to correct for reconstruction, identification, and isolation efficiencies of simulation to data.

4.4   Jets

Most *pp* collision events result is quarks or gluons. Due to quark confinement within QCD, these partons cannot exist independently and undergo a process called hadronization, resulting in colorless cone-like sprays of hadrons. The cones, or jets, are reconstructed using tracks from the ID and energy depositions with the HCAL. The energy deposition and correlated tracks undergo the same topo-cluster algorithm defined for electrons that was discussed in section 4.2. Two types of jets are defined depending on the radii of the cone. There are small-radius (SR) jets and large-radius (LR) jets. Figure 21 shows two jets deposited in the blue colored HCAL, one from a proton, the other from a neutron.

4.4.1   Jet Definition

Since quarks cannot exist independently due to quark confinement, jet reconstruction must rely on complex algorithms exploiting jet kinematics in order to identify the initial quark. The jet reconstruction algorithm must allow for reliable comparison to theory and experiment [77]. In order to identify jets associated to the hard-scatter event, the jet algorithm must be insensitive to soft (low momentum) and collinear (low angle) hadrons.

The current ATLAS standard of such algorithms is called the anti-k_T algorithm [77]. This process identifies hadronic energy depositions as inputs and forms jets through sets of criteria. First it chooses some object i in a list of objects, calculates the distance d_{ij} between two objects within the list and then it calculates the distance d_{iB} between object i and the beam (B). This is shown in Eq. 4.2 .

$$\mathrm{d}_{ij} = \min(p_{\mathrm{T},i}^{2p}, p_{\mathrm{T},j}^{2p})\frac{\Delta_{i,j}^2}{\mathrm{R}^2}\, ,$$

$$\mathrm{d}_{iB} = p_{ti}^{2\mathrm{p}}$$

(4.2)

Where:

$$\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

(4.3)

and $p_{\mathrm{T},i}$, y_i and $\boldsymbol{\phi}_i$ are respectively the momentum, rapidity, and azimuthal angle of object i. The radius of the jet cone is denoted as R and the parameter p is to help govern relative power of the energy. For the anti-k_{T} algorithm the parameter p is set to -1. The algorithm then finds the minimum between d_{ij} and d_{iB}. if the minimum is the distance between the two objects i and j, then the four momenta of the objects are combined. Whereas, if the d_{iB} is the minimum, object i is regarded as a jet. This procedure continues as long as objects remain on the list.

Due to the inverse dependence on p_{T}, soft objects are combined with close-by hard objects (high momentum) before two soft objects are combined [77]. Jets tend to have a conical shape around the hardest object i when using anti-k_{T} algorithm for reconstruction. However, when two hard objects are within a relative distance of $\Delta(i,j) \leq 2R$ of one another, it's difficult to differentiate the jet substructure of either jet. If the two hard objects are within $R < \Delta(i,j) \leq 2R$ then the anti-k_{T} algorithm forms two jets. The shapes of these two hard object jets depends on the transverse momenta of both, ending in one shape to be conical where the other is partially cone-shaped. If two hard objects are found within $\Delta(i,j) < R$, a single jet is formed and the shape is defined by the momenta of the two hard objects. If the momenta are similar, the jet is can be constructed of two cones (with radii lower than R) inside one larger cone. Figure 23 shows several jets classified using the anti-k_{T} algorithm, along with a large-radius jet containing two small-radius jets.

Two types of jets are presented in this thesis. Both are the previously stated SR and LR jets. The cone radii for these types correspond to R = 0.4 for SR and R = 1.0 for LR. More about these two types are explained in the following sections.
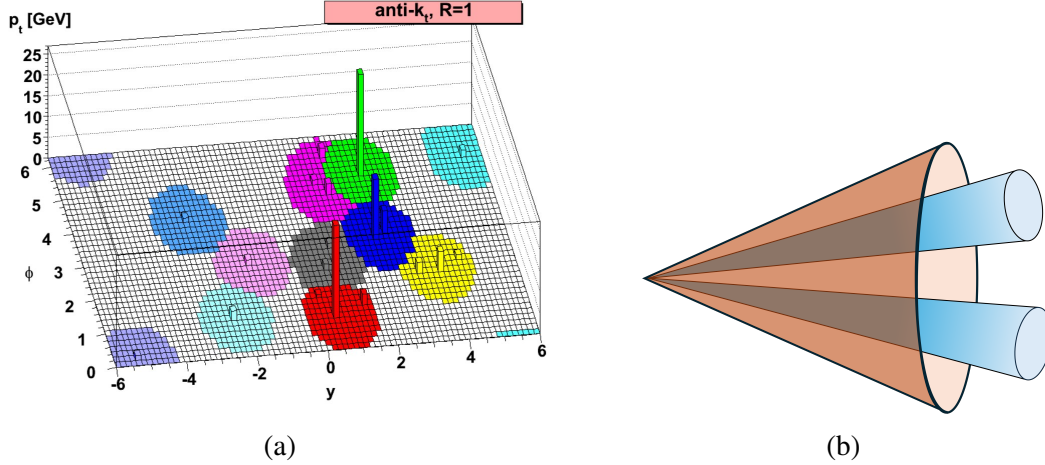
Figure 23: (a) Shapes and the substructures of jets formed using the anti-k_T algorithm in the $\boldsymbol{\phi}$-y plane [77]. The height in the z-axis corresponds to the momentum of the hard objects. Figure (b) Substructure of a large-radius jet containing two small-radius jets of similar p_T

4.4.2 Small Radius Jets

There are several approaches to defining a SR jet. The jets used in this thesis, and are currently ATLAS standard objects, are called *particle flow* or *PFlow*. PFlow objects are the building blocks of PFlow jets. These objects are clusters within the calorimeters and their constituted tracks [23, 25]. Proto-clusters as previously defined in Section 4.2 are standard PFlow objects [43]. The energy depositions in the ECAL and HCAL used to form these proto-clusters are calibrated to the EM scale in order to get a correct measurement of the showers [25]. A weighted mean of the cluster cells is calculated. This value is used to assign $\boldsymbol{\phi}$ and η coordinates, as well as the energy deposited [27]. From here, it's assumed this cluster points to the origin of the coordinate system of the detector. The proto-cluster is taken as a particle of zero mass and has its four-momentum calculated [27]. All the weighted clusters are assumed to be produced from objects originating from the hard-scatter, an origin correction is then applied [25], altering the momentum values to point towards the primary vertex.

SR are defined as these corrected calorimeter proto-clusters with their associated tracks. A previously used jet class called *EMTopo* jets was used. But this class only utilized the proto-cluster energy depositions and didn't use the track information from the ID. The ID has very precise mo-

mentum resolution for low momentum tracks, combining these to the calorimeter energy deposi-
tions greatly increased jet energy resolution in low energy jets. Giving jets improvements in its
energy and direction while also lessening dependence on the number of pile-up interactions [25].

Tracks defined as PFlow objects but adhere to a tighter criteria than previous cases. A crucial
energy subtraction procedure is performed to prevent double counting of energy contributions of
charged particles leaving tracks in the ID and energy deposited in the calorimeters [23, 25]. Energy
that can be associated to a PFlow track is subtracted out of the clusters. A distance metric is applied
between the center of the cluster and its possibly associated track as defined in Eq. 4.4.

$$\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2} \tag{4.4}$$

Where σ_ϕ and σ_η represent the angular topo-cluster width, calculated as the standard deviation
of the displacements, ϕ and η, of the cluster cell center [23]. Taking the smallest $\Delta R'$ to be matched
to a track succeeds in virtually all particles with $p_T > 5$ GeV. If no topo-cluster is found in a
cone with $\Delta R' = 1.64$, it is assumed the particle did not form a cluster in the calorimeter and no
momentum procedure is applied [23].

For each track matched to a cluster, the energy deposited by the particle is evaluated using
simulated events. After a charged particle traversed the ID, it is possible that it would leave more
than one topo-cluster in the calorimeters. If this is the case and the topo-cluster has energy below
the expected amount, clusters within $\Delta R' = 0.2$ of the track are combined. Once the track-to-cluster
matching occurs, energy is subtracted from the calorimeter topo-clusters. Energy deposits around
the subtracted clusters that are found within reasonable shower fluctuations are also subtracted. The
final PFlow object is considered the total subtracted energy from the calorimeter and the matched
tracks that are compatible to the primary vertex.

4.4.3   Large Radius Jets

As the momentum of massive particles (e.g. W, Z, Higgs, or the top-quark) increases, they become
more Lorentz boosted. When these particles decay, their decay products are also highly Lorentz

boosted, collimating them in a single direction. It is advantageous, in this case, to adjust the radius parameter R from the small-R jets and to increase it in order to contain all the collimated decay products. The parameter for large-R jets is increased to $R = 1.0$. By increasing the radius parameter, multi-pronged jet substructure from two-body or three-body decays is much more effectively captured [26].

The reconstruction of LR jets is complicated due to the presence of soft radiation, i.e. energy deposition from underlying pile-up and uncorrelated jets. These degrade the reconstruction performance, LR jet mass resolution and other sub-structure quantities. LR jets are typically reconstructed using the anti-k_{T} algorithm and historically have been based solely on calorimeter energy measurements which have provided excellent energy resolution. By only using energy depositions in the calorimeters, it is difficult to reconstruct separate particles within the LR jet radius of $R = 1.0$, especially where the calorimeter resolution is coarse. It is important to distinguish the LR jet sub-structure and therefore several PFlow algorithms are implemented to extract this information. A variant of the PFlow algorithm called Track-CaloClusters (TCC) was designed to reconstruct jet sub-structure even at the highest transverse momenta. The structures identified using TCCs and PFlow is called Unified Flow Objects, or UFOs. Pile-up mitigation techniques are also implemented such as constituent subtraction, Voroni subtraction, SoftKiller, and pile-up per particle identification (PUPPI).

A trimming procedure is enacted inside the LR jet to reduce the sub-structure constituents that may be related to soft radiation. Several criteria are implemented in this trimming procedure, such as using the ratio of the p_{T} of the sub-jet constituents. This trimming procedure uses the k_{t} algorithm to create sub-jets inside the large radius, creating jet objects with a radius parameter R_{sub}. Any sub-jets with $p_{\text{T}_i}/p_{\text{T}}^{\text{jet}} < \text{f}_{\text{cut}}$. Where p_{T_i} is the transverse momentum of the ith sub-jet, and f_{cut} is the cut parameter. The optimized parameters used for this trimming procedure are $\text{f}_{\text{cut}} = 0.3$ and $R_{\text{sub}} = 0.2$. Figure 24 shows a diagram of a trimmed jet.

Mass is then assigned to each trimmed jet according to their four-momentum squared. To improve trimmed jet mass resolution for high-p_{T} LR jets, an additional mass (m_{TA}) is added based on
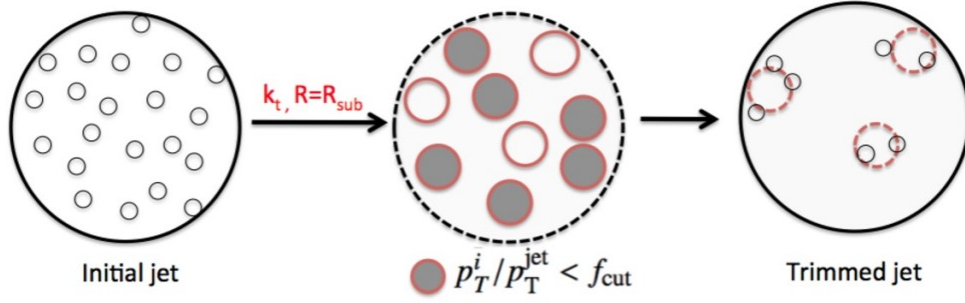
Figure 24: Diagram illustrating the LR jet trimming procedure [21].

tracks associated with the jet, calculated using Eq. 4.5. Tracks are associated to sub-jets through a technique called ghost-association. A pseudo-particle is reconstructed for each track, having the same coordinates and direction as the track. These pseudo-particles are input into the anti-k_{T} algorithm along with the energy clusters in the calorimeter. These pseudo-particles are very soft and therefore do not alter the resulting sub-jets. A track is associated to a sub-jet if a corresponding pseudo-particle is clustered into the sub-jet. This technique of using ghost-association allows for irregular shaped jets, resulting in an occasional non-cone shaped jet.

$$m_{\mathrm{TA}} = \frac{p_{\mathrm{T}}^{\mathrm{jet}}}{p_{\mathrm{T}}^{\mathrm{track}}} m_{\mathrm{track}} \tag{4.5}$$

Where $p_{\mathrm{T}}^{\mathrm{jet}}$ is the transverse momentum of the LR jet, $p_{\mathrm{T}}^{\mathrm{track}}$ is the transverse momenta of a trimmed anti-k_{T} LR jet formed from the tracks associated to the jet and m_{track} is the mass of the track jet. The m_{track} is multiplied by this ratio to include energy from the track of the sub-jet. To further improve the jet mass resolution, a combined mass is calculated from a weighted average of the two masses as shown in Eq. 4.6

$$m_{\mathrm{comb}} = a \cdot m_{\mathrm{jet}} + b \cdot m_{\mathrm{TA}} \tag{4.6}$$

The weights of a and b must satisfy $a + b = 1$ and are derived from resolutions of m_{jet} and m_{TA}. The final calculated value of m_{comb} is the mass of the LR jet.

Energy and mass for the trimmed LR jets are calibrated in data and simulation to equal values

found at particle level in simulated di-jet events [21]. These calibrations are referred to as jet energy scale (JES) and jet mass scale (JMS). The JES is corrected first and subsequently the JMS. In-situ calibrations are taken to account for imperfect detector simulations are derived for JES, JER, and JMS [24]. These are performed on all three values of mass from Eq. 4.6, m_{jet}, m_{TA} and m_{comb}

4.4.4   PFlow Object Calibration

The built PFlow objects are used as input into the anti-k_{T} algorithm as discussed in Section 4.4.1 [77] with $R = 0.4$ to define SR jets. These jets undergo a multi-step calibration. The energy of a reconstructed jet is corrected to match the reconstructed jet at particle level [25]. This step is known as jet energy scale (JES).

Pile-up particles produced in pp collisions can deposit energy that may alter the kinematics of the defined SR jet. These contributions are subtracted in the first step of the JES calibration. The energy is compared to di-jet events at particle level from simulation, correcting the four-momentum vectors [25]. The next step is referred to as the global sequential calibration (GSC). This step corrects dependent parameters such as particles initiating a jet, shower fluctuations, and possible showers not detected by the calorimeters or caught in the muon spectrometer. These corrections are evaluated again by using simulated di-jet events. The final step is to find additional correction factors to mitigate remaining discrepancies to jets from data. The corrections are called in-situ JES calibration factors and are obtained by evaluating events that contain jets and well calibrated objects, such as electrons, muons or photons. Using momentum balance between these objects, correction factors and calculated and applied [25]. Systematic uncertainties associated with these correction factors arise from modeling in the simulation

The jet energy resolution (JER) is also calibrated. The resolution of an object in the detector depends on background noise of the electronics and energy deposits from pile-up. The corrections are found comparing the JER from simulated di-jet events to that of data. The correction factors are derived from these di-jet events and are applied by smearing the energy distribution to match that of data. Systematic uncertainties are obtained in this calibration step.

4.4.5 Jet Vertex Tagger

Taggers are tools developed by the ATLAS Flavor Tagging (FTAG) group to help identify which hadron flavor reconstructed jets originate from. The jet vertex tagger (JVT) is a tagger used to help identify if tracks are associated to the primary vertex of the hard-scatter [6]. There are many *pp* collision that occur in a bunch crossing which is the reason for pile-up. So to correctly identify tracks of the collision of interest is crucial. The JVT is a multivariate tagger algorithm that uses the jet p_T and its associated tracks. It uses the likelihood of the track being associated to a different collision vertex as a discriminating variable. This tagger is only used on jets within $|\eta| < 2.5$, for any object that are more forward, a different algorithm is used called the forward jet vertex tagger (fJVT) [7].

The JVT and fJVT efficiencies are measured in data and simulation [6, 7]. These are determined through simulated events of $Z \rightarrow a +$ jets, employing a tag and probe method. Scale factors are derived to correct for discrepancies in the efficiency measurements of data and simulation.

4.5 Heavy-Flavor Tagging

Identifying heavy flavors of quarks plays a crucial role in particle physics analyses. Many interesting physics analyses have b- or c-quarks in their final state. Such as the SH $\rightarrow b\bar{b}b\bar{b}$ final state which is discussed in the final part of this thesis. Flavor tagging (FTAG) helps discriminate background processes from signal events. Therefore, this process has an important role in not just finding new physics, but also precision measurements. Since quarks undergo hadronization and cannot exist independently, the quarks themselves cannot be tagged, but rather a color neutral states, b- and c-hadrons. The b-hadrons have a lifetime of around 1.5 ps [53], thus decaying after traversing the length of 2.5 mm when it has about 30 GeV. The b-hadron is also associated with a high decay multiplicity of an average of five charged particles. Figure 25 (a) shows the charged particle decay count from a b-hadron (B^0), where (b) shows the p_T fraction the b-hadron carries within the jet which peaks at about 70%. This results in another important property of a b-jet, it has a high semi-

leptonic decay fraction. The typical b-jet looks like Figure 26. The b-hadron has a decay length L and is indicated by the red dotted line. This leads to a secondary vertex. The decay products from the secondary vertex may lead to a tertiary vertex, especially if it's a c-hadron. The displacement of the secondary vertex with respect to the primary vertex is given by a parameter called the decay length. The displacement of a track with respect to the primary vertex is parameterized with impact parameters (IPs) denoted as d_0 in the figure. This d_0 IP tends to be large for b-hadrons due to the long decay lifetime and large mass (5 GeV).
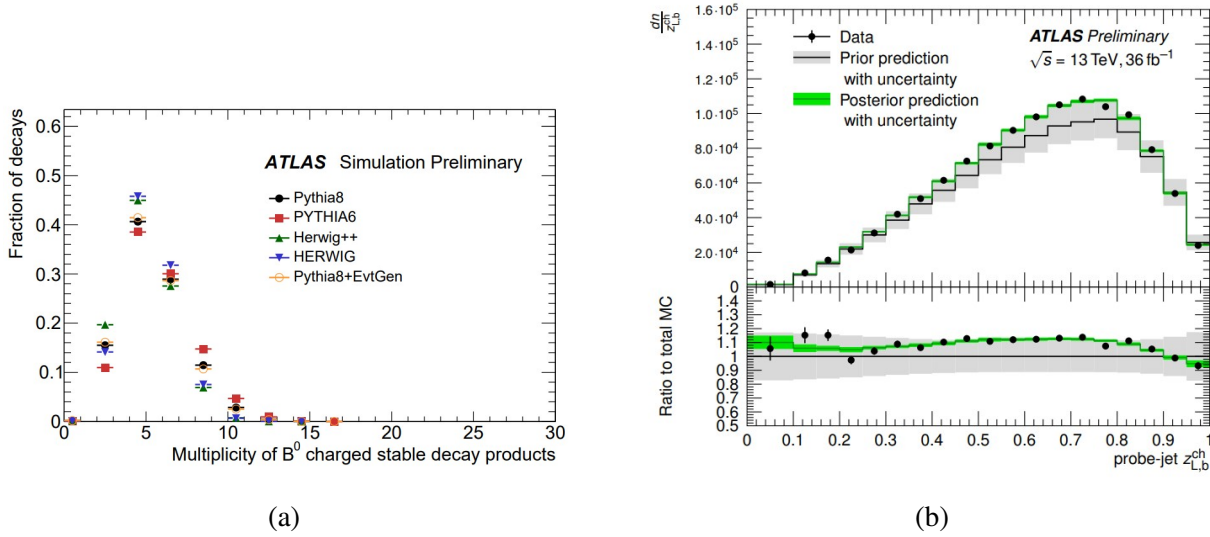
Figure 25: (a) Decay multiplicity of the b-hadron d_0 into stable charged products with compared MC generators. (b) Fragmentation of b-hadron p_T

C-hadrons tend to have less stable charged particle multiplicity, shorter decay lengths (resulting in a smaller d_0) and are lighter. This results in similar jet topologies, but not identical. Light-jets originate from lighter quarks which results in tracks being associated with the hadronization itself. These properties of c- and light-jets make them separable from b-jets. All these unique b-jet properties are targeted by FTAG tools developed to *tag* a b-hadron. There are several b-tagging algorithms, all attempting to extract specific features of the b-hadron's information. These baseline algorithms can be categorized into three categories: the IP based algorithms called Impact parameter 2(3) Dimensional (IP2D, IP3D), Recurrent Neural Network IP (RNNIP) and Deep Impact Parameters (DIPS); the vertex-based algorithms of Secondary Vertex (SV1) and JetFitter; lastly the
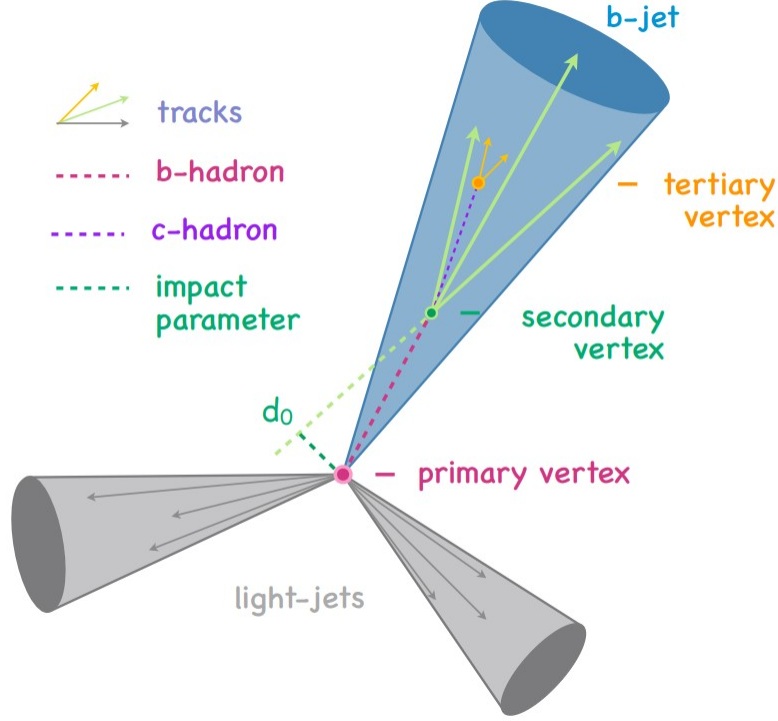
Figure 26: Diagram view of a b-jet with a first, second, and third vertex labeled.

soft-muon tagger (SMT).

These baseline taggers are then combined *high-level taggers* such as a multivariate boosted-decision tree (MV2), deep-learning based DL1 (DL1), and a graph neural network (NN) tagger (GN1). DL1 is studied for the HL-LHC in chapter V and the following sections will describe its related baseline taggers. Figure 27 is a diagram showing the baseline input for each of the high-level taggers.
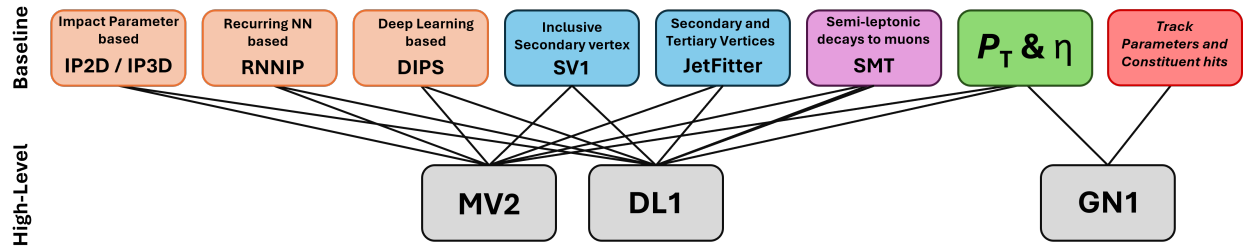
Figure 27: Diagram of FTAG baseline and high-level taggers.

4.5.1 Impact Parameter Algorithms

A b-hadron has the longest lifetime of any quarks which results in a long decay length and a displaced secondary vertex. The IP is the point of closest approach of tracks from the b-hadron decay to the primary vertex. This parameter can be split into two parts, the d_0 as seen in Figure 26, and a longitudinal part $z_0\sin\theta$. A variable of the IP is the signed lifetime significances $s_{d_0} = \frac{d_0}{\sigma_{d_0}}$ and $s_{z_0} = \frac{z_0\sin\theta}{\sigma_{z_0\sin\theta}}$. are calculated corresponding to the IP divided by its uncertainty, as shown in Figure 28. The sign of this variable is assigned by extrapolating the track back to the primary vertex. This value is negative if the jet axis has to be extended backwards from the primary vertex to cross the track or its projection, otherwise it's positive. The IP algorithms are used to satisfy certain criteria such as: $_T^{track} > 1$ GeV, the IPs have to fulfill $|d_0| < 1$ mm, and $|z_0\sin\theta| < 1.5$ mm. Additional requirements are required for the number of hits in the silicon layers: $N^{Si}_{hits} \geq 7$ as well as an upper limit of silicon and pixel layer holes $N^{Si}_{holes} \leq 2$ and $N^{pixel}_{holes} \leq 1$. The following sections briefly describes the IPxD baseline taggers and the DIPS tagger. The RNNIP tagger will not be described as it is outside the scope of this thesis.
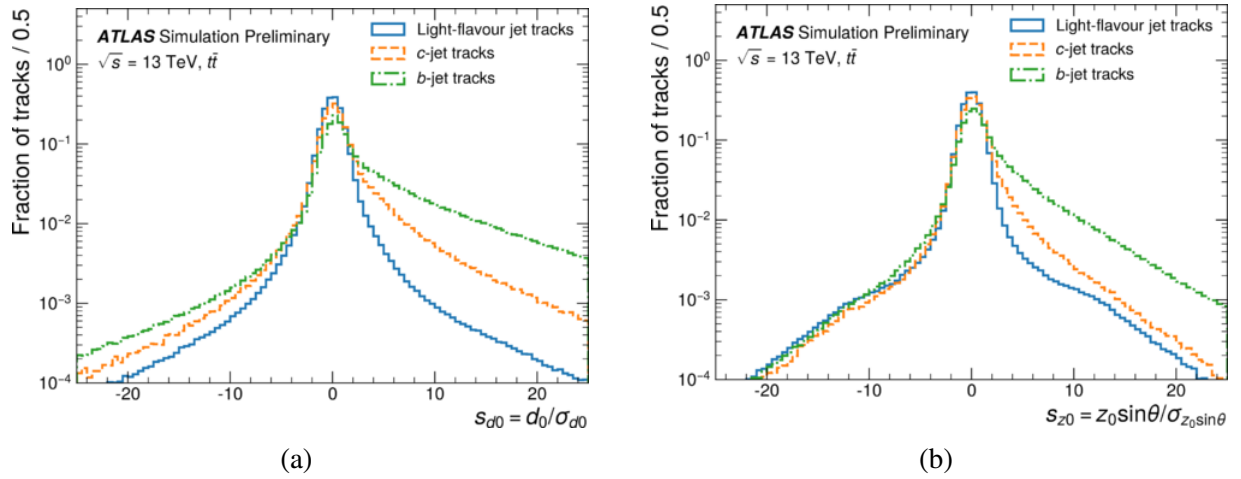
Figure 28: The signed IP significances variable shown for b-, c-, and light-jets for (a) transverse and (b) longitudinal components in $t\bar{t}$ events

IPxD Discriminating Variables	
Variable	Description
	LLR based on signed significance
IPxD$_l$	b- from light flavor jets
IPxD$_c$	b- from c-jets
IPxD$_{cl}$	c- from light flavor jets

Table 6: Variables for the IP2D and IP3D taggers (3 each)

IPxD

The IPxD taggers comprise of two algorithms: the IP2D which only uses the transverse IP d_0 variable which tends to be less sensitive to pile-up, and the IP3D tagger which uses both the transverse d_0 and $z_0 \sin\theta$ variables [8]. The track categorization is based on pixel layer hit patterns as pre-defined by reference templates for b, c, and light hadrons. The final discriminant is a Log-Likelihood ratio (LLR) of probabilities of being any of the three hadrons. This LLR is described by Eq. 4.7.

$$\text{IPxD}_{l,c,cl} = \sum_{i \in \text{tracks}} \log \left(\frac{P^i_{b,b,c}}{P^i_{l,c,l}} \right) \tag{4.7}$$

The probability density functions to calculate P_b, P_c and P_l are extracted from MC simulations. Table 6 shows the six different variables for the IPxD tagger used to tag jets. The IPxD baseline taggers are used as input into the high-level taggers such as MV2 and DL1.

DIPS

The DIPS is a machine-learned algorithm that uses the IP variables as input for a Deep Sets architecture [54], treating the elements as a set without any specific order. The formalism of DIPS has p_i which is the vector representation of the inputs associated with the ith track in the jet, then the Deep Sets architecture applies its weight ϕ to each track. Tracks are then summed over and has additional processing in the form of a feed forward NN (F) as described in Eq. 4.8.

$$\mathcal{O}(p_1, ..., p_n) = F \left(\sum_{i=1}^{n} \phi(p_i) \right) \tag{4.8}$$

where $\mathcal{O}(p_1, ..., p_n)$ represents the b-, c-, and light- class probabilities. The architecture bisects the problem into operations over the inputs and over the sets. The track network ϕ extracts the relevant track information and the forwarding jet-network F accounts for the correlations between the tracks [11]. Permutation invariance of the set is encoded using the sum operation. Since DIPS encodes this permutation invariance, a much more natural representation of the data is made, allowing the machine learning (ML) algorithm to be trained more effectively. This algorithm outperforms IPxD and RNNIP algorithms.

The ML model of DIPS outputs class probabilities of b-, c- and light-jets (p_b, p_c, p_l). A discriminant is made from a combination of these three probabilities dictating whether a jet is b-tagged or not, this discriminant is given by Eq. 4.9.

$$D_b = \log\frac{p_b}{(1 - f_c)p_l + f_c p_c} \tag{4.9}$$

where f_c is a free parameter that helps balance the rejection rate between light-jets and c-jets for a given b-tagging efficiency. This is optimized post-training. A few input variables for DIPS are scaled and shifted due to a mean that is not close to zero, this process is described in reference [11]. A list of DIPS input features can be seen in Table 7. The features that required additional preprocessing are noted.

4.5.2 Secondary Vertex Algorithm

The secondary vertex SV1 algorithm reconstructs a single displaced vertex in a jet using tracks. Due to hardware constraints, sometimes the track resolution in the ID is not always able to resolve the entire decay cascade of a hadron in every jet. Therefore the criteria of reconstruction is only a single vertex, which is a good approximation for a b-jet. The first step matches all two-track vertices, rejecting vertices that are compatible with tracks associated with long-lived particles, photon conversions or hadronic interactions with detector material. The next step is combine each accepted two-tracks into secondary vertices while removing all rejected tracks. Important information can

DIPS Input Features		
Input	Description	Preprocessed
s_{d_0}	d_0/σ_{d0}: Transverse IP significance	
s_{z_0}	$z_0\sin\theta/\sigma_{z0\sin\theta}$: Longitudinal IP significance	
$\log p_T^{frac}$	$\log p_T^{frac}/p_T^{jet}$: Logarithm of fraction of the jet p_T carried by the track	✓
$\log\Delta R$	Logarithm of opening angle between the track and the jet axis	✓
IBL hits	Number of hits in the IBL: could be 0, 1, or 2	
PIX1 hits	Number of hits in the next-to-innermost pixel layer: could be 0, 1. or 2	
shared IBL hits	shared IBL hits	
split IBL hits	Number of split hits in the IBL	✓
nPixHits	Combined number of hits in the pixel layers	
shared pixel hits	Number of shared hits in the pixel layers	
split pixel hits	Number of split hits in the pixel layers	
nSCTHits	Combined number of hits in the SCT layers	✓
shared SCT hits	Number of shared hits in the SCT layers	

Table 7: Input variables for DIPS

be extracted from SV1 such as the vertex mass, decay length and its significance, number of associated tracks as well as the ΔR between the jet and the secondary vertex. Table 8 shows an overview of the variables for SV1.

SV1 Variables	
Variable	Description
$N_{trkAtVtx}^{SV1}$	Number of tracks associated to the SV
$N_{2trkAtVtx}^{SV1}$	Number of reconstructed two-track vertices candidates within the jet
m_{inv}^{SV1}	Invariant mass of the SV calculated from the associated tracks
f_E^{SV1}	Energy fraction of the SV associated tracks with respect to all tracks of the jet
$\Delta R(jet, SV)$	ΔR between the jet axis and the secondary vertex relative to the primary vertex
L_{xy}^{SV1}	Reconstructed SV transverse decay length
L_{xyz}^{SV1}	Reconstructed SV decay Length
S_{xyz}^{SV1}	Decay Length significance

Table 8: Variable overview for the SV1 algorithm [8]

4.5.3 JetFitter

The decay chain multi-vertex reconstruction algorithm, a.k.a. JetFitter, is the second displaced vertex finder algorithm behind SV1 [8]. This algorithm aims to reconstruct the decay cascade topology of weakly decaying b- and c- hadrons. It assumes the primary, secondary, and the tertiary vertices are aligned in one line in the direction of the hadron's trajectory. The goal of this assumption is to cope with the finite track resolution. After an initial track selection of removing tracks associated to the primary vertex using a Kalman Filter [60]. The resulting variables of JetFitter are shown in Table 9.

This algorithm allows for special variables for the c-hadron exploiting the fact that the JetFitter vertex would be close to the primary vertex. These are chosen to differentiate the topologies between the b- and c- hadron. C-hadrons have lower decay multiplicity than the b-hadron due to their lower mass, thus decay products carry a larger momentum percentage along with a larger rapidity with respect to the jet axis. These variables are shown in Table 10.

JetFitter Variables	
Variable	Description
m_{inv}^{JF}	Invariant mass of tracks associated to one or more displaced vertices
f_E^{JF}	Charged jet energy fraction in the secondary vertex
S_{xyz}^{JF}	Decay length significance of the displaced vertex
$N_{1\text{-trk vertices}}^{JF}$	Number of 1-track displaced vertices
$N_{\leq 2\text{-trk vertices}}^{JF}$	Number of vertices with more than one track
$\Delta R^{JF}(P_{jet}, P_{vtx})$	ΔR between jet axis and the vectorial sum of all track momenta
N_{trks}^{JF}	Number of tracks associated to SV
$N_{vertices}^{JF}$	Number of reconstructed vertices

Table 9: Variable overview of JetFitter algorithm for b-tagging [8]

JetFitter C-Hadron Variables	
Variable	Description
L_{xyz}^{JF}	Displacement of SV from the primary vertex
L_{xy}^{JF}	Transverse displacement of SV from the primary vertex
$min(Y_{trk}^{JF})$	Minimal rapidity of tracks within the jet
$max(Y_{trk}^{JF})$	Maximum rapidity of tracks within the jet
$avg(Y_{trk}^{JF})$	Average rapidity of tracks within the jet
$min(Y_{trk, SV}^{JF})$	Minimal rapidity of SV tracks
$max(Y_{trk, SV}^{JF})$	Maximum rapidity of SV tracks
$avg(Y_{trk, SV}^{JF})$	Average rapidity of SV tracks
m_{inv}^{JF}	Invariant mass of tracks associated to the SV
ΔE^{JF}	Energy of tracks associated to SV
f_E^{JF}	Charged jet energy fraction of SV
N_{trks}^{JF}	Number of tracks associated to the SV

Table 10: Variable overview of JetFitter algorithm for c-tagging [8]

4.6 High-Level Taggers

Baseline taggers are optimized for single properties of b-hadron decays. High-level taggers are multivariate taggers comprised of the baseline taggers previously described and as seen in Figure 27. There are three high-level taggers that are currently deployed in ATLAS: the boosted decision tree (BDT) based tagger MV2, the deep learning based DL1 and the graph NN GN1. Currently, DL1 is the standard tagger within ATLAS, but GNN based taggers have proven to be more effective and are currently being developed to replace it. DL1 is explored in higher detail in chapter V.

4.6.1 Working Points

As briefly discussed in Section 4.3, WPs are deployed for the high-level taggers to cover various needs in physics analyses. These WPs are defined using the b-tagging efficiency on $t\bar{t}$ MC samples. These single cut WPs are used since having a full continuous b-tagging discriminant would require a continuous calibration in extremely fine efficiency bins. This task would take enormous time and require immense complexity. The tagging efficiency is described in Eq. 4.10

$$\varepsilon^{j} = \frac{N_{pass}^{j}(\boldsymbol{D} > T_f)}{N_{total}^{j}} \tag{4.10}$$

where $N_{pass}^{j}(\boldsymbol{D} > T_f)$ are the number of jets of flavor j passed the cut T_f on the tagger discriminant $\boldsymbol{D}$ and N_{total}^{j} are all jets of flavor j before the cut. The WPs are defined based on b-jet efficiency ε^{j} evaluated on $t\bar{t}$ MC samples. In ATLAS there are four WPs defined as seen in Table 11. The inverse of each WP characterizes the c- and light- jet rejection rate. Misidentification improves with lower signal efficiency, therefore rejecting more background, trading off lower signal statistics. Within a physics analysis, every jet that passes a WP is considered a b-jet. Each high-level tagger that is described in the following sections have different rates of b-tagging efficiencies, meaning the number of tagged b-jets in an analysis will depend on which tagger is used. Therefore high tagging efficiencies are sought after, encouraging physicists to use the latest technological advances. The BDT based MV2 was the standard tagger until Deep Sets architectures overtook and shown to be much more efficient, thus DL1 has now become the new standard. Recent developments in the past year has shown that graph NNs can exploit track information much more profoundly, resulting in much higher tagging efficiencies than the DL1 tagger and is expected to take its place.

4.6.2 MV2

The high-level tagger MV2 used to be the recommended tagger for EMTopo jets for Run 2 of ATLAS. This tagger used 24 inputs from the baseline taggers as input for the BDT training as well as p_T and $|\eta|$. Training the MV2 tagger used what is called a *hybrid* sample which is a mixture of $t\bar{t}$

Working Points for B-Tagging in ATLAS	
Cut	b-tagging efficiency
loose	85%
medium	77%
tight	70%
very tight	60%

Table 11: Summary of b-tagging single cut WPs

and Z' events. This is to ensure there is a large coverage of the p_T spectrum. The BDT used b-jets as the signal class and c- and light- jets as a background class. To balance the performance between the c-jet and light-jet rejection, a c-jet fraction f_c was set to 7% and therefore a light jet fraction of 93%. The end result was named MV2c10. Figure 29 shows the performance of the MV2, DL1 and baseline taggers in terms of background rejection as a function of b-tagging efficiency.

(a)   (b)
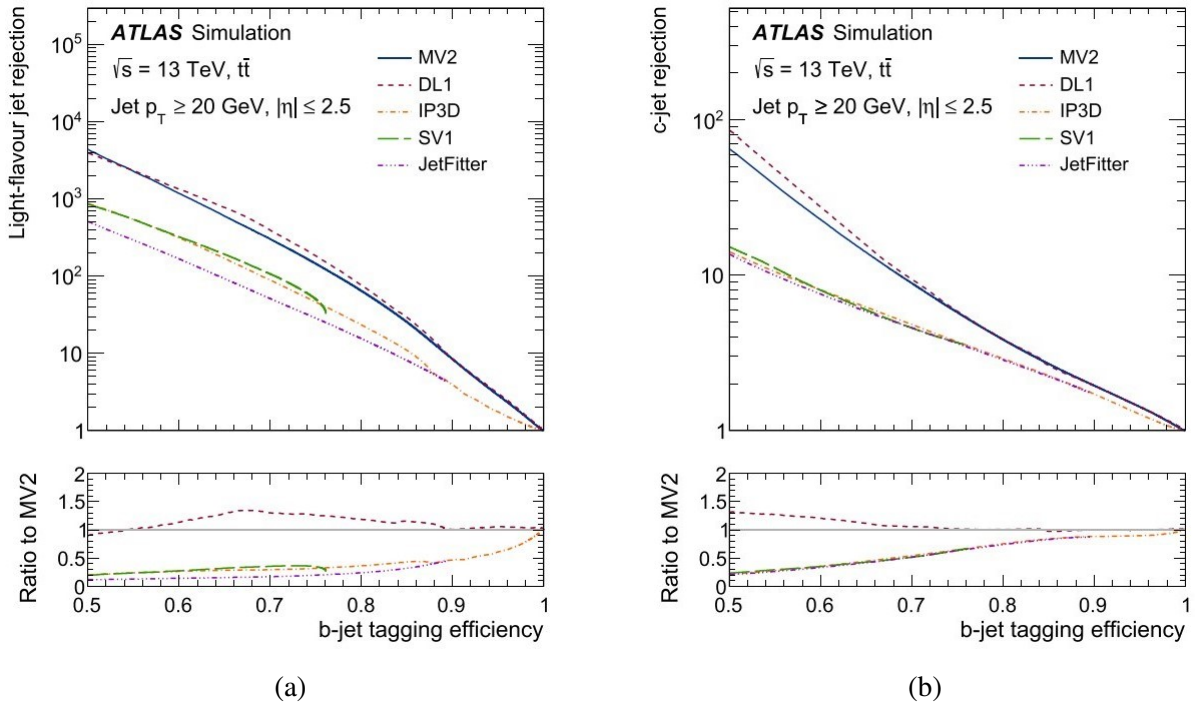
Figure 29: The (a) light-jet and (b) c-jet rejections versus the b-jet tagging efficiency for the baseline taggers IP3D, SV1, JetFitter and the high level taggers MV2 and DL1. Evaluated on $t\bar{t}$ MC events [10]

4.6.3 DL1

The high-level DL1 tagger is a deep feed-forward NN with three output nodes that correspond to the b-, c-, and light- flavor jet probabilities. The ReLU activation function is used for each layer while the last layer uses the softmax activation function. Using the softmax function for the output layer allows the output to be interpreted as probabilities. The final b-tagging log-likelihood discriminant score is calculated from the multi-class output using Eq. 4.11.

$$\mathcal{D}_b(f_c) = \log \left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_l} \right) \qquad (4.11)$$

where p_b, p_c and p_l are the output scores of DL1 representing the probabilities for the jet to be a b-jet, c-jet or a light-jet, respectively. The c-jet fraction f_c allows tuning to balance the performance of the c-jet and light-jet rejection. The c-jet rejection increases as a function of f_c and light-flavor decreases as a function of f_c. An advantage of using the DL1 compared to MV2 is that the discriminant can be rewritten to perform c-jet tagging with b-jet rejection. This new rewritten discriminant can be seen in Eq. 4.12

$$\mathcal{D}_c(f_b) = \log \left(\frac{p_c}{f_b \cdot p_b + (1 - f_b) \cdot p_l} \right) \qquad (4.12)$$

where f_b is a floating value for the b-jet fraction. DL1 has shown many positives over MV2. The tagger is effectively maintained with much less person power, while also fewer variables have to be calculated and stored, saving computing power. The DL1 tagger is a family of high-level multivariate taggers. These taggers differ from the baseline tagger inputs, they are as follows: baseline DL1, DL1r, DL1rmu, and DL1d. The DL1 baseline uses the same variables as MV2 with the additional JetFitter variables for c-jet identification. DL1r configuration uses the RNNIP instead of the baseline IPxD taggers. The DL1rmu exploits the soft-muon information. Lastly, the DL1d uses the baseline taggers SV1, JetFitter but uses the DIPS instead of the baseline IPxD. More information is given on the DL1d tagger in Chapter V.

4.6.4 GN1

The GN1 tagger is currently the latest and greatest particle identification algorithm. It differs from the previously described high-level taggers in a few ways. GN1 does not make use of the underlying baseline taggers but instead exploits the internal structure of the jet through the use of two auxiliary training objectives: the grouping of tracks originating from a common vertex, and the prediction of the underlying physics processes [14]. The graph neural network takes 2 kinematic and 21 track variables as training input. Using auxiliary tasks removes the need to add the baseline taggers, therefore simplifying the training process and allowing the tagger to be more versatile in training phase space. The algorithm is trained on *truth-information* obtained through simulations.

The GN1 combines a graph neural network architecture with auxiliary training objectives. Initially, the two jet inputs (transverse momentum and signed pseudorapidity) are fed into a per-track initialization network with three hidden layers, each containing 64 neurons. These three layers are a Deep Sets architecture such as the technique used by DL1. A fully connected graph is built from the outputs of these three Deep Set layers. Each node h_i in the graph represents a single track in the jet, characterizing a feature vector. The output nodes from the Deep Set architecture are used to populate the graph. Outputs from each graph layer are aggregated features of each node h_i and neighboring nodes N_i. The feature vectors of each node are fed into a fully connected NN layer that produces updated values of the vector. These updated features are used to compute edge scores of each node (scores calculated from neighboring nodes). A non-linear activation function is used while a softmax function calculates weights for each pair of nodes using the edge scores. Finally, an updated node representation is computed by taking the weighted sum over each updated node representation. After the graph network is implemented, these outputs are used in a node classification network to predict track truth origin. A diagram of the entire network architecture is seen in Figure 30. The track and jet kinematic input variables for the GN1 tagger are listed in Table 24 within the Appendix APPENDIX A:. The next version of GN1 is called GN2 and is currently expected to replace its predecessor.
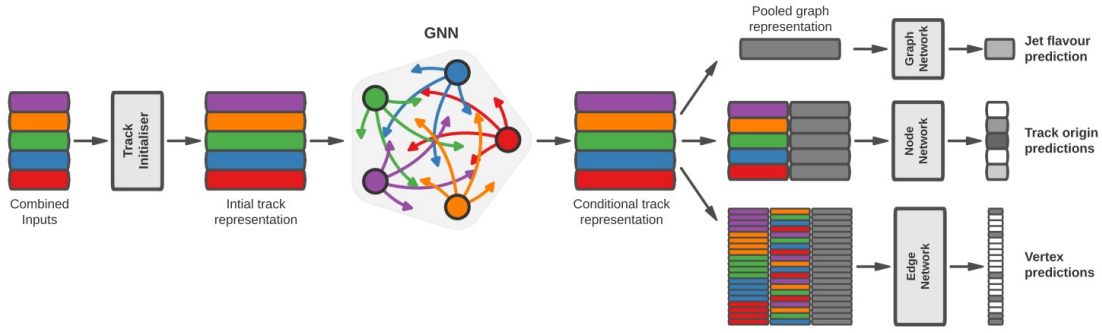
Figure 30: The network diagram of GN1. First, a Deep Sets architecture is used to populate node features in a GNN. The GNN outputs are used to predict jet flavor, track origins, and track-pair vertex compatibility. [14]

4.7   Event Simulation

Simulated *pp* events are used in every physics analysis while also being vital in event reconstruction calibrations. They also play an important role in simulating BSM models which are used to help define new phases spaces for the chance of detecting new physics. Simulated events are used later in this thesis as ML input and are vital in deploying efficient particle identification algorithms. BSM simulations are also used in the Anomaly Detection analysis in Chapter VI to show separation between standard SM events and the anomalous BSM events.

Event simulations not only simulate kinematics of particles within the SM (or beyond of it) but are also used to show how simulated particles would interact with the ATLAS detector. Due to the complexity of integral emerging from QFT calculations, Monte Carlo MC techniques are used in the simulations [91].

4.7.1   Monte Carlo Generators

Every event generation is typically divided into two parts: the matrix element generation that describes the hard scatter and secondly the parton showering and hadronization modeling which includes the initial state radiation (ISR) and the final state radiation (FSR). The matrix element and the parton shower can be calculated mostly perturbatively, other calculations cannot. There are

several common mathematical models used to simulate hadronization: the Lund string model and the cluster model. In the Lund string model, the connection between a quark and an antiquark is modeled as a string, assuming the potential between the two to be linearly increasing with distance. These strings then split according to a Fragmentation function forming a new quark-antiquark pair. This process continues until only stable hadrons remain [17]. The cluster model is based on QCD confinement where neighboring partons build color clusters which then decay into two hadrons who also will decay until stable hadrons are formed [73]. Figure 31 shows a simplified MC simulation.

The steps of calculated the matrix elements, parton showers, hadronization and the underlying soft radiation can all be simulated by the common generators PYTHIA8 [88], HERWIG7 [52], SHERPA [46], POWHEGBOX [66], and MADGRAPH5_@aMCNLO [74]. A few drawbacks to some of these generators are: PYTHIA8 provides mainly leading order calculations which are typically not sufficient since next-to-leading order (NLO) corrections would have to be applied and these can be fairly large. HERWIG7 is only used for parton showering even though it does include NLOs, the fraction of negative event weights can be very large. The generators of POWHEGBOX and MADGRAPH5_@aMCNLO are used to provide higher order calculations while also being able to be interfaced into PYTHIA8 and HERWIG7 to calculate the parton showering. To describe models that use non-perturbative processes, parameters must be tuned using collision data. The most common used parameters in ATLAS are the A14 [22] parameters for PYTHIA8 or the H7UE [50] parameters for HERWIG7.

4.7.2 Detector Simulation

The last step of simulation is to simulate the ATLAS detector. MC generators don't take into account the detector, the final state output needs to be translated to a signal that represents the detector's output. A full ATLAS detector simulation is done in two parts. The first step is to incorporate the detector geometry, this is done by a program called GEANT4 [57], this provides highly precise modeling of particle interactions with each sub-detector. This process is so precise that it takes a large fraction of computing power that ATLAS uses. Therefore techniques have been developed to decrease this computational power by mimicking the output of GEANT4 in the use of fast simulator

algorithms [35]. These algorithms use thousands of individual parametrizations of calorimeter response. This process uses a lot less computational resources with a trade-off of precision. These fast simulations are widely used by ATLAS.
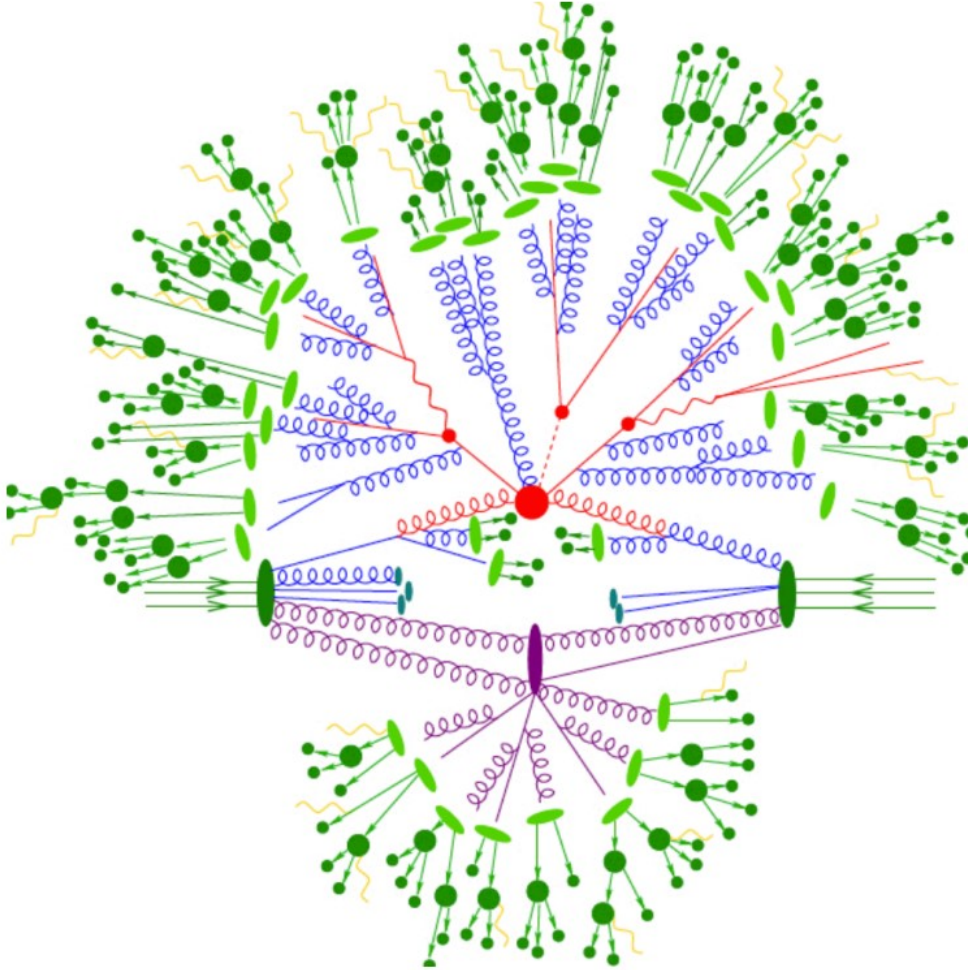
Figure 31: Illustration of a hadron-hadron collision simulated by a MC generator. The center red circle signifies the hard scatter collision while the purple oval represents underlying soft-scatter events. The red and blue tree-like structures depict QCD bremsstrahlung simulated by parton showering. The other elements are hadronization (light green), hadron decays (dark green), and photon radiation (yellow). [70]

CHAPTER V

DEEP LEARNING MODEL TRAINING FOR THE HL-LHC

The high-level tagger DL1 was introduced during Run 2 of the ATLAS detector as the second high-level tagger beside MV2 and originally used *EMPTopo* jets. The tagger was designed to combine the base-line taggers as discussed in 4.5 into a final discriminant score as seen in Eq. 5.1. Introducing the human-designed detector level base-line taggers into a machine learned algorithm brought several benefits such its multi-class output of the probabilities for each flavor of jet (b-jet, c-jet and light-jet). Creating a NN brings flexibility and customization which is why they are preferred over BDTs.

This next chapter discusses the training of the DL1 tagger using the DIPS base-line tagger, creating the DL1d high-level tagger. This model was trained using simulations of the ATLAS detector with its new geometry after its upgrade to transition into Run 4. Implementing the ITk detector as discussed in Chapter 3, Section 3.3. These simulations include the increased pile-up from the HL-LHC of $\langle \mu \rangle = 200$. The jet structure used are the *PFlow* discussed in Section 4.4.2.

5.1 DL1 Design

The underlying structure the the DL1 tagger is a deep feed-forward NN with three output nodes corresponding to b-, c-, and light-flavor probabilities as seen in Figure 32. The ReLU activation function is used for each hidden layer while the output layer utilizes the softmax activation function to ensure the output can be interpreted as probabilities. These multi-class outputs are then used to calculate a final discriminant score as seen in the reintroduced DL1 discriminant log-likelihood

function 5.1.

$$D_b(f_c) = \log\left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_l}\right) \qquad (5.1)$$

Here p_b, p_c, and p_l are the multi-class probability outputs for b-, c-, and light-flavor jets. The f_c is an introduced c-jet fraction tunable parameter which can give emphasis to the c-jet or light-jet rejection performance. This parameter is expected to be tuned for each tagger as its value depends on the needs of the physics analyses. As previously discussed, this fraction can gives rise to the ability to flip the DL1 tagger to focus on b-jet and light-jet rejection, requiring the reintroduced log-likelihood discriminant equation 5.2.

$$D_c(f_b) = \log\left(\frac{p_c}{f_b \cdot p_b + (1 - f_b) \cdot p_l}\right) \qquad (5.2)$$

Here there are the same multi-class probability scores as seen in 5.1 but the tunable parameter has now changed to a b-jet fraction, f_b. Due to its ease of training and maintainability, the DL1 tagger requires less man-power and creates reliable tagging scores that reduce the amount of necessary stored information, saving on memory. As discussed previously, the DL1 family contains four different types of taggers that utilize different baseline information. The first is the baseline *DL1* that uses the same input baseline tagger variables as MV2 (as seen in 6, 8,9) with the additional JetFitter variables for c-tagging as seen in table 10. From there the baseline *DL1* is used with additional NN inputs. *DL1r* uses the baseline *DL1* with additional flavor probabilities from the RNNIP algorithm. *DL1rmu* exploits the soft-muon information from the muon spectrometer while also including the inputs from the *DL1r*. Lastly, the latest tagger is the *DL1d* which is the baseline *dl1* that includes an additional feed-forward NN called DIPS that was previously discussed in Section 4.5.1. A diagram of the DL1 family is seen in Figure 33.

5.1.1   DL1d design

In the following sections, training of the DL1d model for the HL-LHC is described. Since the DL1d utilizes the deep learning model DIPS, its design if fairly more complex than the other DL1
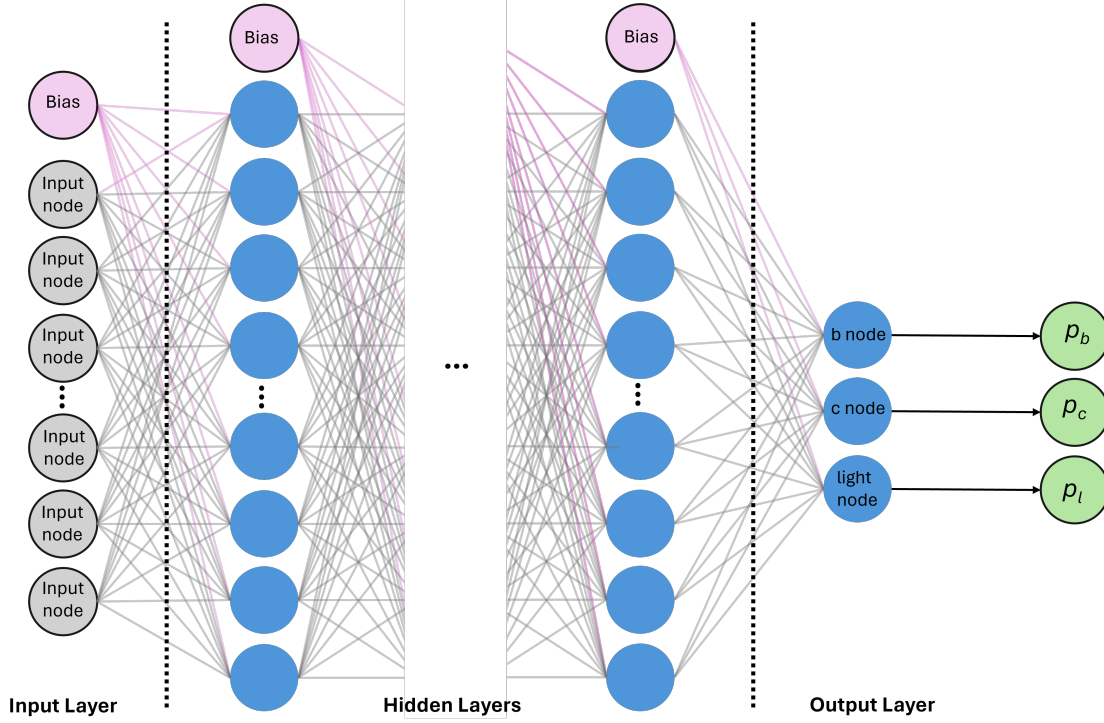
Figure 32: Deep feed-forward architecture of DL1 with unspecified number of nodes

models in the family. The architecture design can be seen in Figure 34. Notice there are two sub-networks ϕ and F which correspond to the DIPS architecture. The DIPS model is trained using track variables described in Table 7. Each track in a jet is first processed through the network ϕ. In the next step, all n track networks, corresponding to the number of tracks in a jet, are summed up and further processed via the network F. This can be summarized via an equation as seen:

$$\vec{P}_i = F\left(\sum_{i=1}^{n} \phi(\vec{x}_i^t)\right) \tag{5.3}$$

where $\vec{x}_i^t$ are the track input features and $\vec{P}_i$ is the vector of the b-, c-, and light flavor jet class probabilities corresponding to the DIPS output.

The DIPS networks, ϕ and F, are joined to the DL1d feed-forward architecture denoted as $\mathcal{U}$. A joint architecture allows the passing of more information into the DL1 NN $\mathcal{U}$. In the case of Figure 40, a layer of 30 nodes is concatenated with the other jet features that are processed through a network containing 72 nodes. A full back-propagation up to all the track NNs ϕ is done during the
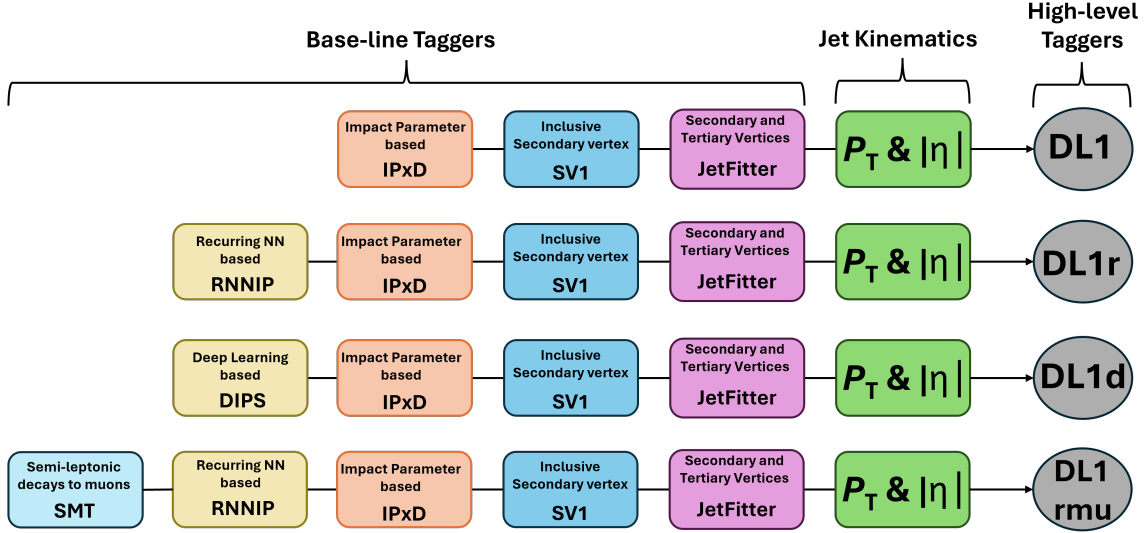
Figure 33: Structures of the DL1 tagger family, differing in NN input variables

training allowing a joint optimization of all three networks. The DIPS networks have intermediate losses which are used to determine the performance and optimization of the DIPS model prior to being used as input to the DL1 network. The final feed-forward DL1 architecture $\mathscr{U}$ has three output nodes to calculate the multi-class probabilities. Both networks have dedicated losses. While the loss of the final $\mathscr{U}$ network $L(\mathscr{U})$ is sensitive to track features and jet kinematics, the loss of F network $L(F)$ is only sensitive to tracks. The overall optimization is performed over the combined loss, as shown:

$$L(\text{comb.}) = L(\mathscr{U}) + \lambda \cdot L(F) \tag{5.4}$$

5.2   Software Chain

The software chain from training the DL1 tagger to validation and deployment in the ATLAS software `ATHENA` are unassociated. The training of the NN is based on the industry-standard open-sourced coding language Python3.6 [85]. The packages numpy [44] and pandas [82] are used for data handling while it's formatted in the HDF5 format [65]. The ATLAS software package developed to convert root datasets into HDF5 format is called the Training Dataset Dumper (TDD) The code structure is configured using human-readable formats such as JSON [48] and yaml to
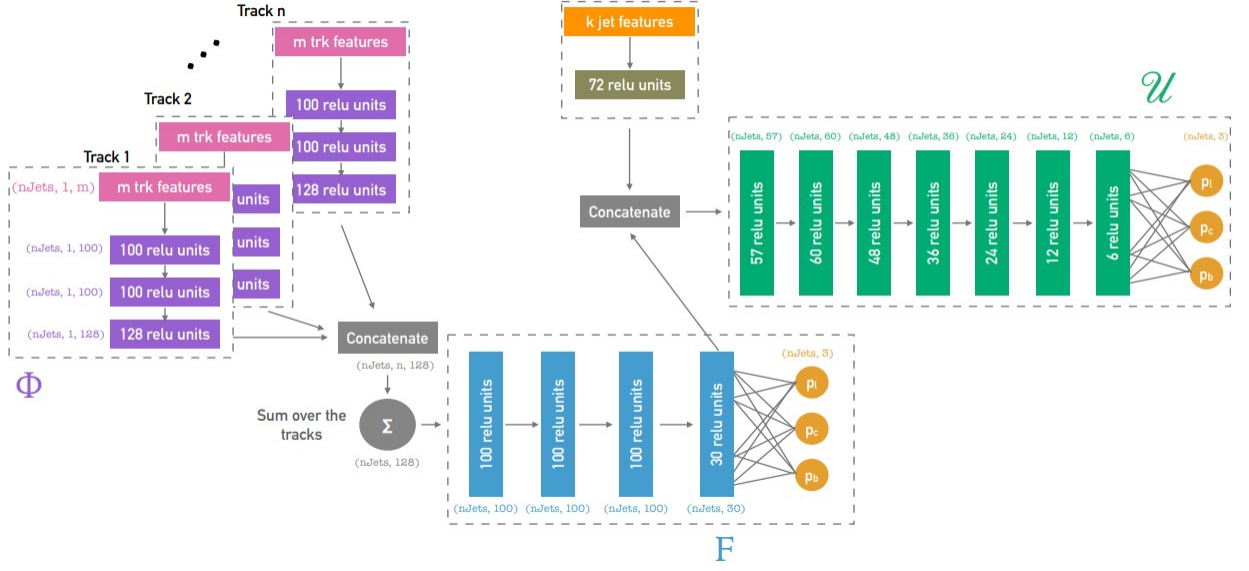
Figure 34: Architecture of DL1d combining the DIPS φ and F networks to the DL1 feed-forward nodes 𝒰

ensure user compatibility. For the NN training, Tensorflow [55] is used with a Keras2 [49] frontend. These packages are used to create a user-friendly framework for deep learning training called Umami which handles the preprocessing and training steps. For visualization, the package matplotlib [69] and tools from scikit-learn [47] are included in the training scripts. These tools are compiled into a convenient tool in the Umami family named Puma. The full workflow (Umami and Puma) can be performed using a Docker image [80] which allows the software to be ran from any computer without the need to install the necessary environment. The output model is then transformed into a format called LightWeight Tagger Neural Network [45] (lwtnn) which integrates the model into the ATLAS software ATHENA.

5.3 Preprocessing

The first step before training a DL1 model is preparing the input in a step called *preprocessing*. To ensure robustness of a model to perform properly over a large p_T spectrum, two simulated samples are combined into a *hybrid sample*. One sample is of $t\bar{t}$ events and the other is of a BSM particle

called Z′. An example p_T spectrum of these two samples can be seen in Figure 35 in (a). The Z′ allows a flat p_T spectrum of up to 4.5 TeV with a total range up to 6 TeV. This Z′ sample is known as an *extended Z′* sample since the normal distribution of a Z′ is up to about ≈4 TeV. The $t\bar{t}$ sample populates the lower p_T spectrum, ensuring there are enough statistics for a proper model to be trained while the Z′ populates larger p_T to increase tagging efficiencies for heavier particles. This sample, along with all that is included in this thesis, are using PFlow objects. The object selection within the $t\bar{t}$ sample is of the following:

$$t\bar{t} \text{ selection} : \begin{cases} \text{b-jets} & \text{b-hadron}_T < 250 \text{ GeV} \\ \text{c-jets} & \text{jet } p_T < 250 \text{ GeV} \\ \text{light-jets} & \text{jet } p_T < 250 \text{ GeV} \end{cases}$$

As noted here, the b-jet selection uses b-hadron p_T. The sample fraction of $t\bar{t}$ and Z′ can be adjusted, in the following example plots, the fraction is adjusted to 70% $t\bar{t}$ and 30% Z′. The merged *hybrid sample* can be seen in (b) of Figure 35.

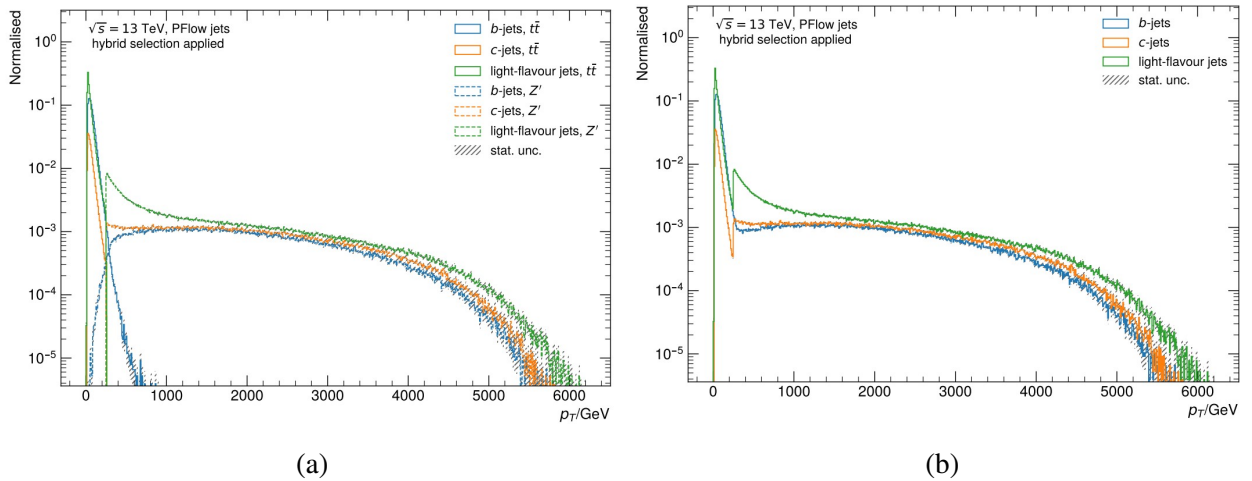Figure 35: (a) p_T distribution of $t\bar{t}$ sample (solid lines) and a Z′ sample (dashed lines). The $t\bar{t}$ b-jet distribution is normalized to unity and all other distributions are normalized to the b-jet distribution. (b) Samples from (a) merged into one distribution normalized to b-jet distribution

The goal of the DL1 tagger is to differentiate between hadron flavors. As seen in Figure 36, the

flavor composition is imbalanced towards light-flavor jets. In order to avoid kinematic and count biases within the tagger, a resampling step must be applied. This also helps mitigate discontinuity between the two combined samples. Kinematic correlations are important for the baseline taggers, the goal for the high-level tagger is to avoid differences in kinematics and attempt to tag hadrons by jet-by-jet classification. Instead of weighting each sample fraction to match a chosen distribution, a resampling method is deployed to avoid instabilities in the model training steps. There are two methods of resampling. There is *undersampling* where single jets are removed from the majority classes to fit a distribution of a minority class or there is *oversampling* where jets from minority classes are duplicated to match a given distribution. Figure 37 shows a diagram of both methods. A combined method called Importance Sampling with Replacement, also referred to as *PDF* sampling, undersamples and oversamples distributions to match a middle target distribution while enforcing the same shape. As seen in Figure 36, b-jets are a middle distribution, therefore the PDF sampling is used in the preprocessing step of the following work.
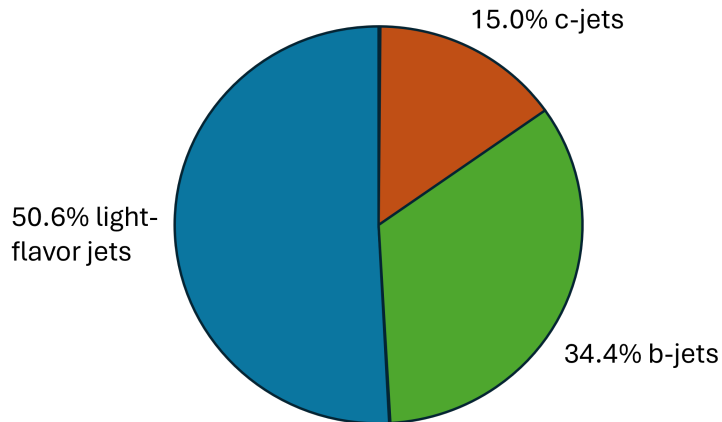
Figure 36: Hybrid sample hadron fractions.

The obvious issue that is present in these sampling methods is the duplication of events during oversampling a minority distribution. This may cause a bias due to the lack of diversity of statistics in the minority groups. Luckily, the samples are MC produced and therefore the statistics can be simply increased by producing more samples through event generators. However, a resampling procedure referred to as the *count method* was studied as a comparison to the PDF method. The count method strictly only undersamples to the lowest distribution, thus drastically decreasing statistics in
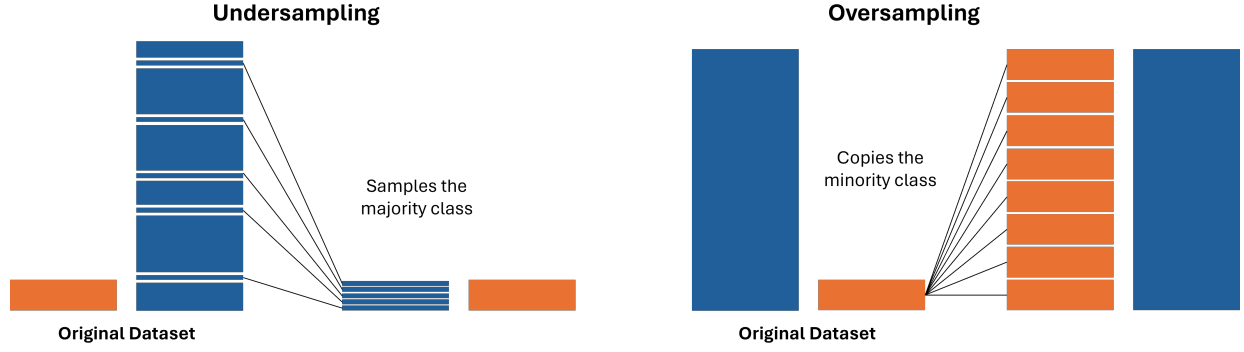
Figure 37: Diagram of the resampling methods to ensure balanced model training.

favor of bias possibilities. The p_T and $|\eta|$ b-jet distributions are taken as reference and the c-jet and light-jet flavor are resampled to match them. The binning used for the undersampling procedure tends to be more granular, especially in the lower p_T region where the sample transition bins are, while at higher p_T the bins are wider. This can be seen by resampling the example distribution from Figure 35 in Figure 38.
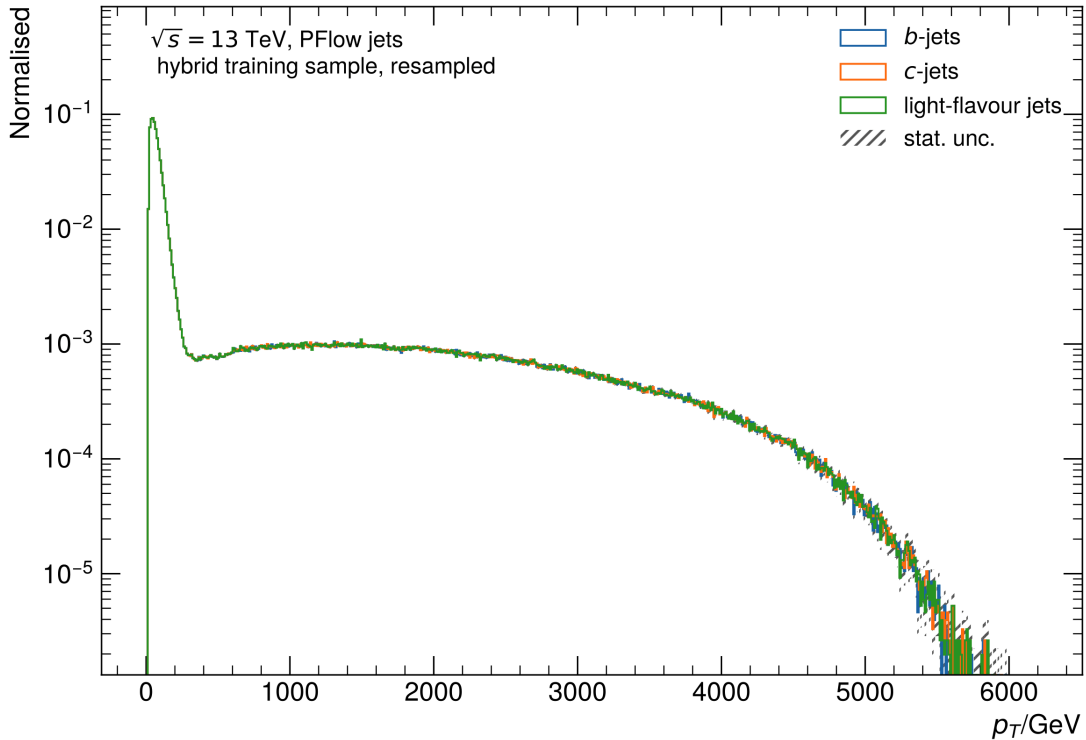
Figure 38: Resampled example flavor distributions.

Lastly before training, the ranges of the input variables need to be balanced so that they're

88

all in the same order of magnitude, otherwise certain variables would carry more weight in the model training procedure. Therefore, all the variables are shifted to a mean of zero and a standard deviation of one, with the exception of binary check variables. Also, extremely anomalous events originating from obscure phase spaces within the samples are removed to secure an undisturbed training process.

5.4 HL-LHC Samples

The current versions of these heavy-flavor taggers have been trained using simulated samples from Run 2. The latest versions that are being implemented as of 2023 have been optimized for Run 3 and are trained on the latest Run 3 samples. In the scope of this thesis, preliminary trainings for the DL1 tagger have been conducted using samples simulated using the ITk from Run 4 that starts in 2029 in order to see how well the new geometry can increase tagging efficiency with the main focus on the increase of $|\eta| \leq 4$. All objects simulated in these samples are PFlow objects as discussed in Section 4.4.2.

5.4.1 Object Selections

The most important features for training effective taggers are tracks, therefore it's vital to simulate tracks through the new ITk detector with precision. These tracks represent trajectories of charged particles reconstructed from hits in the ITk pixels and strip sensors. The η-dependent variables underwent a combination of selections based on general track quality criteria and requirements specific to flavor tagging algorithms. The variables used in the baseline taggers (found in tables 6, 8, 9) have adjusted criteria due to the pseudorapidity extension of the ITk detector. Only jets with $p_T > 20$ GeV and $|\eta| \leq 4$ are considered. A generator-level overlap removal with electrons and muons from W boson decays are also applied. Pile-up jet rejection algorithms are still under development for the HL-LHC, so the selected jets are required to be matched within $\Delta R > 0.3$ with truth-jets built from hard-scatter stable particles that are clustered with the anti-k_T algorithm with R=0.4. The selected reconstructed primary vertex is required to be within 1 mm of the hard

scatter vertex. Tracks are associated to jets using p_T-dependent matching criteria, with a cone size of $\Delta R \approx 0.45$ for jets with $p_T = 20$ GeV to $\Delta R \approx 0.25$ for jets with $p_T > 200$ Gev. In cases where multiple tracks pass the criteria, the closest track is chosen. Jet flavor labels are assigned based on generator-level presence of b- or c-hadrons, or τ hadronic decays. If a b-hadron with $p_T > 5$ GeV is found within $\Delta R = 0.3$, the jet is labeled a b-jet. This procedure is repeated for c-hadrons and τ hadronic decay products. Jets that have no labels by the end of this iteration are assumed to originate from light-flavor quarks and are labeled light-jets. Table 12 shows the new requirements imposed on the jets due to the implementation of the ITk detector.

New Flavor Tagging Requirements for ITk			
Requirements	Pseudorapidity interval		
	$\|\eta\| < 2$	$2.0 < \|\eta\| < 2.6$	$2.6 < \|\eta\| < 4.0$
pixel + strip hits	≥ 9	≥ 8	≥ 7
pixel hits	≥ 1	≥ 1	≥ 1
pixel + strip holes	≤ 2	≤ 2	≤ 2
p_T [MeV]	> 900	> 500	> 500
$\|d_0\|$ [mm]	≤ 2.0	≤ 2.0	≤ 3.5
$\|z_0 \sin\theta\|$ [mm]	≤ 5.0	≤ 5.0	≤ 5.0

Table 12: Training samples used for DL1d for HL-LHC studies [39]

5.4.2 Sample Preprocessing

The following samples consist of the necessary $t\bar{t}$ and Z' samples for training with a CME of 14 TeV which aligns with the predicted HL-LHC beam energy. A small sample of only 100,000 events is also used that include thes HGTD (briefly discussed in Section 3.3.3) detector within the simulation step and is used for a few studies that are mentioned later in this chapter. These three samples are listed in table 13.

The values corresponding to the letters in the latter half of the container names correspond to certain production versions. Table 14 shows the nomenclature for these values.

HL-LHC DL1 Training Samples				
Sample	Container	$\langle \mu \rangle$	# of Events
$t\bar{t}$	mc15_14TeV.600012.PhPy8EG_A14_ttbar_hdamp258p75 _nonallhad.recon.AOD.e8185_s3654_s3657_r12573	200	4,374,000
Z′	mc15_14TeV.800030.Py8EG_A14NNPDF23LO_flatpT _Zprime_Extended.recon.AOD.e8185_s3654_s3657_r12574	200	1,605,205
HGTD	mc15_14TeV.600012.PhPy8EG_A14_ttbar_hdamp258p 75_nonallhad.recon.AOD.e8185_s3770_s3773_r13618	200	100,000

Table 13: Training samples used for DL1d for HL-LHC studies

Sample Production Nomenclature		
Prod. Step Name	Prod. Step Tag	Definition
evgen	e	Event generation tag. MC generation of quark-gluon interactions to parton showering and hadronization
simul	s	Simulation tag corresponds to simulated detector hits when using MC simulations
digit	d	Digitization tag is turning the simulated energy deposits into detector response that resembles raw data
recon	r	Reconstruction tag corresponds to the offline reconstruction algorithm release used to build physics objects
deriv	p	Derivation p-tag refers to a derivation framework used to obtain useful physics objects

Table 14: Training samples used for DL1d for HL-LHC studies

In order to finalize the preparation of these samples, they had to be prepared and resampled due to their imbalance of hadron flavors. The imbalance in both of the upgrade samples are shown in Figure 39. In order to prepare these samples, both of them were combined into a hybrid sample consisting of 70% $t\bar{t}$ and 30% Z′ events. Next, the dataset is split into three different sets. The first is the training set, this one is the largest of the three in order to maximize training statistics. The other two sets are the validation set used to validate the model and lastly is the testing set which tests

the model afterwards. The resampling process of PDF, as previously described in the last section, was used to match the b-hadron distribution whereas the count method undersamples to the lowest distribution which happens to be c-jets. Table 15 breaks down the sample statistics for each dataset and resampling method.
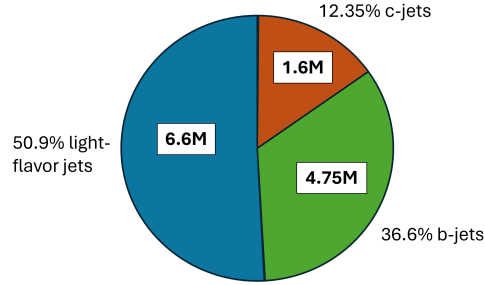

Figure 39: Diagram of hadron composition in the HL-LHC MC samples.

Dataset Compositions and Statistics			
Dataset	Total Types of Jets	From: $t\bar{t}$	From: Z'
Training set	4.75M b-jets	4.1M	650K
	1.6M c-jets	900K	700K
	6.6M light-jets	5.1M	1.1M
Validation set		2.6M	1.1M
Testing sets		2.6M	1.1M
PDF Resampling Method		Count Resampling Method	
15M Training Jets		4.8M Training Jets	

Table 15: Dataset statistics used for training DL1d for the HL-LHC

5.4.3 Training DL1d

The training of the DL1d tagger for the HL-LHC is only preliminary where the final version will have more of a dedicated optimization and increased statistics. The DIPS tagger that is used as input for this preliminary version of DL1d for Upgrade was trained by a fellow graduate student and was

92

trained using two electron selections. Within ATLAS, four fixed values of the electron discriminant are defined (similar to the WP defined for b-tagging). These WPs for the electron selection cuts are referred to as *VeryLoose*, *Loose*, *Medium*, *Tight* cuts. These cuts are set by the efficiencies for identifying a prompt electron with E_T = 40 GeV are 93%, 88%, 80% for *Loose*, *Medium*, *Tight* respectively [37]. One DIPS model contains a tighter WP selection cut, and the other a loose WP selection cut. Therefore, four DL1d models are trained in total, two models trained using both of these cuts and being resampled via the count method and the other two using the PDF method. The DL1d architecture follows the deep feed-forward architecture shown in Figure 32 but with specified nodes. A diagram showing the specified node architecture can be seen in Figure 40.
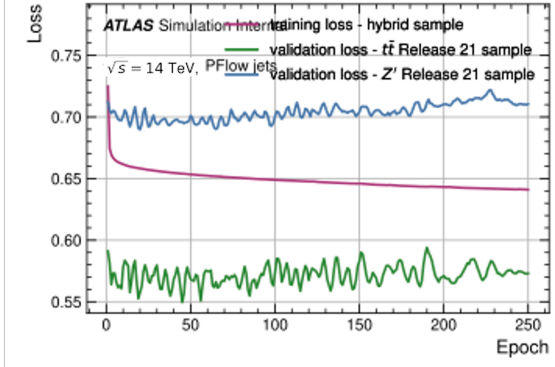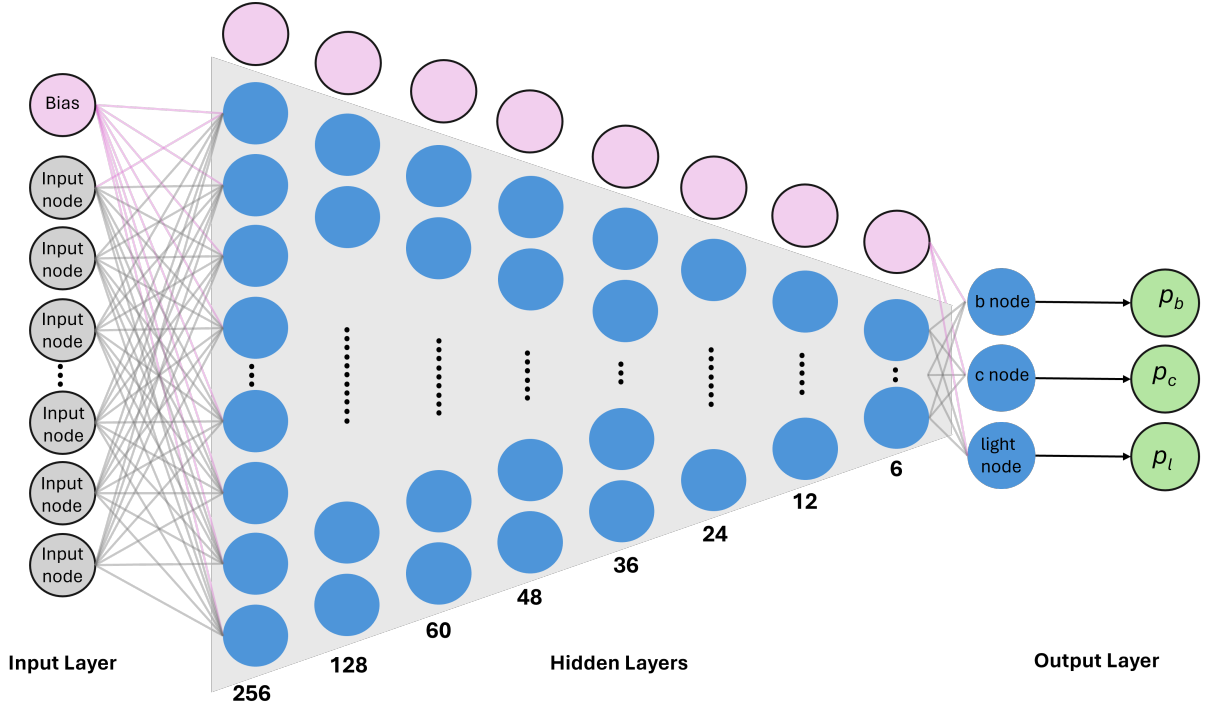

Figure 40: Deep feed-forward architecture used for preliminary Upgrade DL1d model.

Using the loss function described in Eq. 5.4, the loss value is recorded per training epoch and plotted in Figure 41 while also showing the loss of the associated DIPS model used in the DL1d architecture. The loss is shown for the $t\bar{t}$ sample, Z' sample and the combined hybrid sample. Since the training is done on the hybrid sample, the loss function converges much quicker than just on the single samples. Figures 41a and Figure 41b shows the loss per epoch for both the PDF

method and Count method, revealing that loss convergence stabilizes faster in the PDF method than the count. This is due to the increase of statistics, letting the model training being able to reliably find underlying patterns at a faster rate. Figures 41c and 41d shows the light-jet rejection rate with respect to the training epoch. This reveals the increase of effectiveness between the DIPS model and the DL1d. Adding jet kinematic information from the baseline taggers to the underlying track structures from DIPS massively increases light-jet rejection rate. Using the PDF resampling increases this rejection rate as seen between plots (c) and (d) in Figure 41. The convergence in the undersampling method Count takes about 100 more epochs, substantiating that the PDF resampling method is the superior approach. All these figures used the 77% WP for b-tagging.
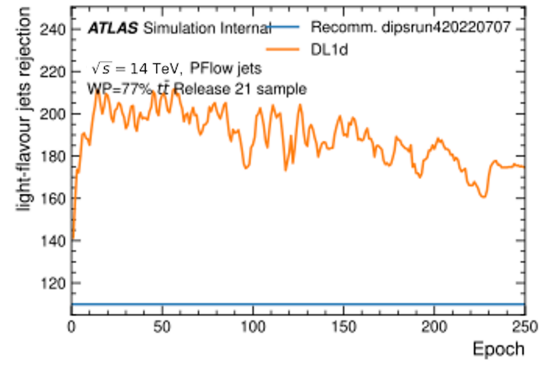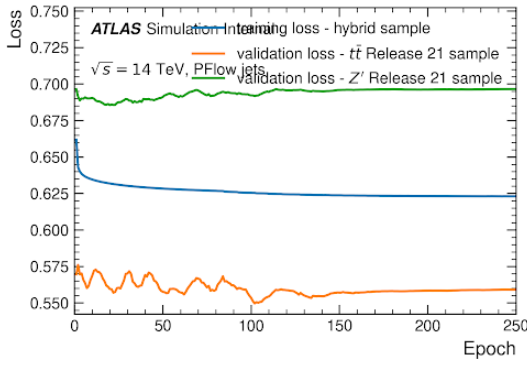
(a) Loss per training epoch using PDF

(b) Loss per training epoch using Count
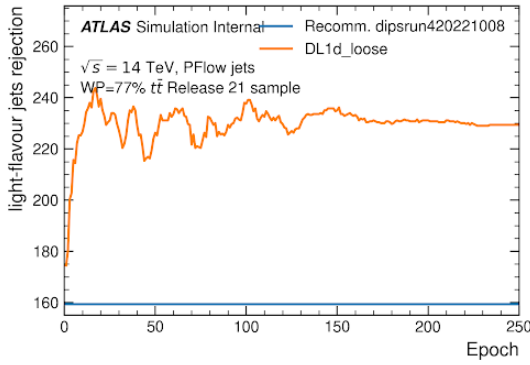
(c) PDF light-jet rejection rate per epoch

(d) Count light-jet rejection rate per epoch

Figure 41: (a) Shows loss per training epoch using the PDF method for both DL1d and its associated DIPS model using the loose electron selection cut. (b) Shows loss per training epoch using the count resampling method for both DL1d and its associated DIPS model using the loose electron selection cut. (c) Shows the light-jet rejection rate per training epoch using the PDF method. DL1d shows rejection at a higher efficiency. (d) Shows the light-jet rejection rate per training epoch using the Count resampling method. DL1d outperform DIPS but does not has a lower rejection rate than using the PDF resampling method as seen in (c)

In order to calculate the final discriminant scores, a fraction scan is implemented to find an effective c-jet fraction value for the float f_c shown in Eq. 5.2. Since light-jet rejection and c-jet rejection are both affected by the chosen floating value, it's optimized to have a balanced effect in the $t\bar{t}$ sample while favoring c-jet rejection in the Z' sample. The fractions are taken at the 77% WP. The scan can be seen in Figure 42. Table 16 shows the actual percentage value of c-hadrons within each sample. The value of 9% was chosen for f_c and can be seen in Figure 42 at the point marked by the red *X*.

C Hadron Percentage	
Dataset	percentage
$t\bar{t}$	7%
Z$'$	9%

Table 16: C-hadron percentages in both training samples prior to combining
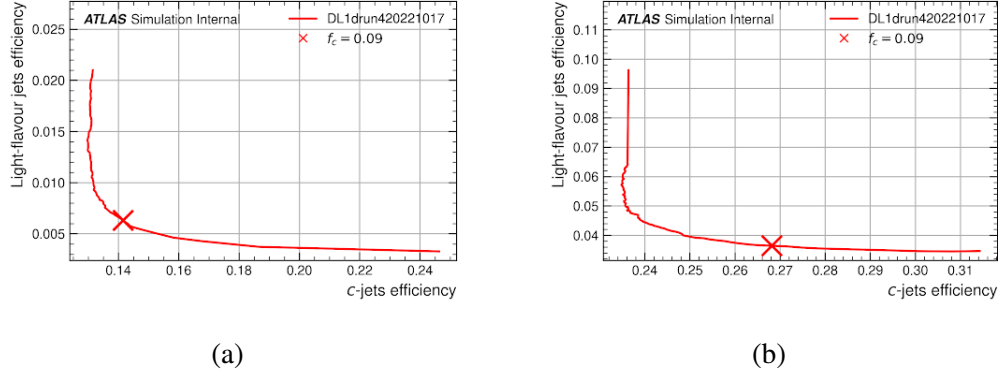

(a)   (b)

Figure 42: C-jet fraction scans for the float value f_c. The balancing value chosen is marked by *X* on both plots. The chosen value is 0.09, balancing the light-jet rejection and c-jet rejection in the $t\bar{t}$ sample while favoring c-jet rejection in the Z$'$ sample.

Since four models were trained using both resampling methods (PDF and Count) while using two different DIPS models that use different electron cuts, they had to be compared to see which model outperforms the other three. While holding the optimized value of $f_c = 0.09$, the b-jet tagging efficiency was checked for all the models. Figure 43 shows a Receiver Operating Characteristic (ROC) curve. This shows background rejection vs b-tagging efficiency. The goal for these models is to have the highest rate of b-tagging efficiency while rejecting the most amount of background. The bottom two plots below the ROC curve are two ratio plots showing the efficiency of the DL1d models with respect to the DL1d using the Count resampling method. The left plot validates the models on the $t\bar{t}$ sample while the right validates them on the Z$'$ sample. It's quite noticable on the $t\bar{t}$ sample plot that using the loose electron cut dramatically increases the DL1d performance b-tagging efficiency. As seen in pink on the right plots in Figure 43, the model DL1d + the loose electron cut DIPS model outperforms the other three while being comparable in the $t\bar{t}$ plot on the

96

left. Therefore, the DL1d model with the loose electron selection cut was taken as the best model and was used in the following comparison plots for the rest of this section.

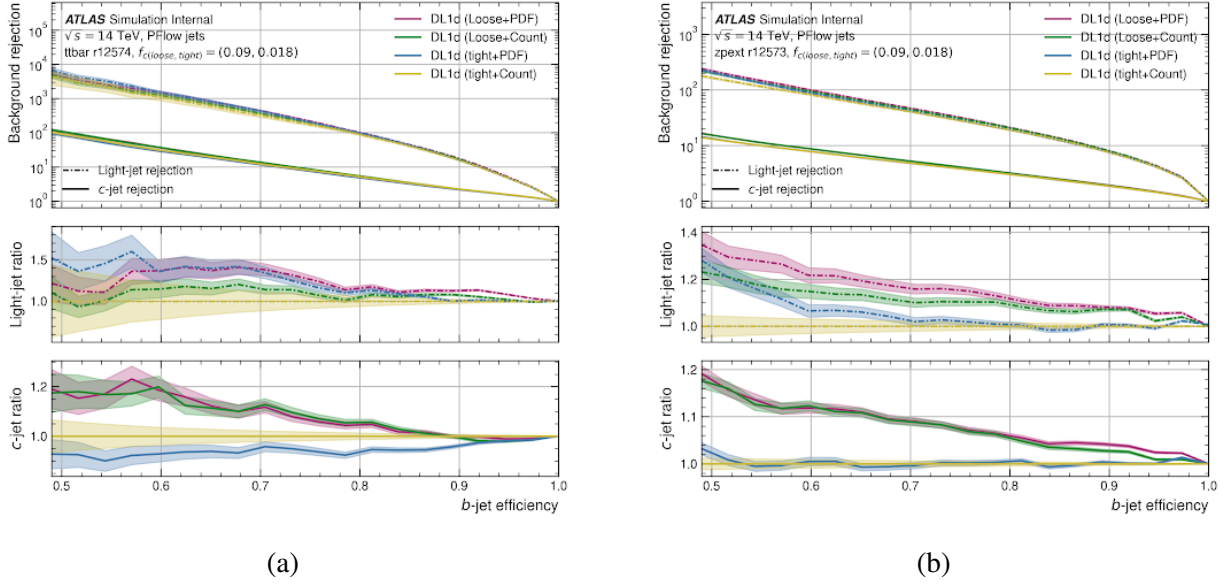(a)   (b)

Figure 43: Comparison plots of all four DL1d trained models. Plot (a) shows the performance of each model using the $t\bar{t}$ sample. Both loose electron WPs outperform the tight WP. Plot (b) shows all four DL1d models validated on the Z' sample. Again, both loose electron WPs models outperform the tight WPs while the PDF resampling method (pink) outperforms the Count method (green). The loose DL1d model using the PDF resampling method was chosen to be the superior trained model.

Once the optimum model was chosen, the next step was to check on how the performance of it compares to the baseline DIPS model. It is expected for the DL1d model to greatly outperform the DIPS model simply due to the fact that the jet kinematics are added after the training of DIPS, increasing the underlying pattern recognition from the excess of available information. This performance can be seen in Figure 44. As expected, the DL1d model outperforms the baseline DIPS model by approximately 25% in c-jet rejection vs b-tagging efficiency.

Just as in testing the model's b-tagging efficiency, the c-tagging efficiency can also be checked using the discriminant seen in Eq. 5.2. The following plots show the performance of both the DL1d and the baseline DIPS tagger for c-tagging efficiency. The DL1d tagger outperforms the DIPS as expected just as in the b-tagging study. For these plots, the HGTD sample was used as listed in Table 13. Two performances are shown for two different floating b-jet fractions f_b as seen in Eq.
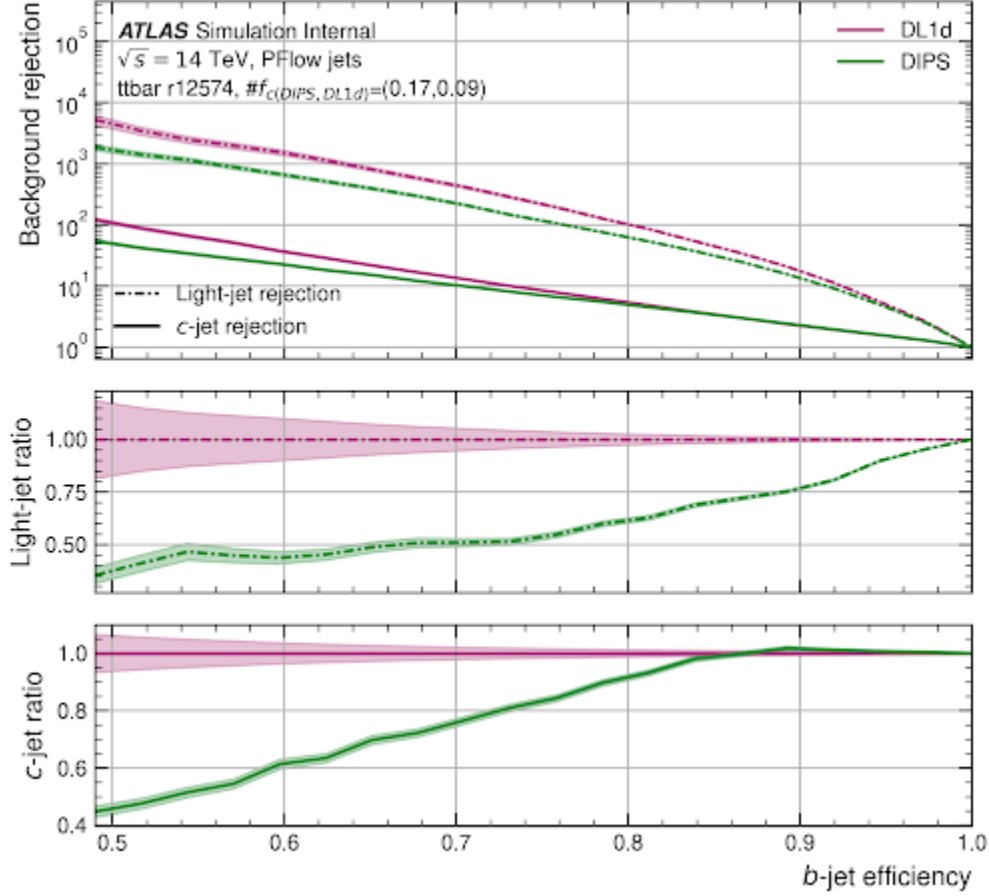
Figure 44: Comparison plot for the DL1d model and the baseline DIPS. DL1d outperform DIPS as expected. The c-jet fraction for the DIPS model was taken to be $f_c = 0.17$ whereas this fraction was $f_c = 0.09$ as previously stated.

5.2. The left plot has a floating b-jet value of f_b=0.24 and the right plot has a value of f_b=0.45.

This DL1d tagger was the first of its kind to be trained on samples with the geometry of ATLAS in Run 4. With the increase of luminosity within the HL-LHC and the upgrades in timing resolution and granularity within the ITk and HGTD, tagging efficiencies is expected to increase. It has proved to be difficult to train a preliminary tagger to outperform the current state-of-the-art taggers. Though, at the current state, this is to be expected for several reasons. One would expect the tagging efficiency to increase with the amount of available statistics to train on. This preliminary DL1d using Upgrade samples had a total of $\approx$15 million jets to train on that include jets between the newly included eta range of 2.4 < $|\eta|$ < 4.0 (total range of 0 < $|\eta|$ < 4.0) which may not
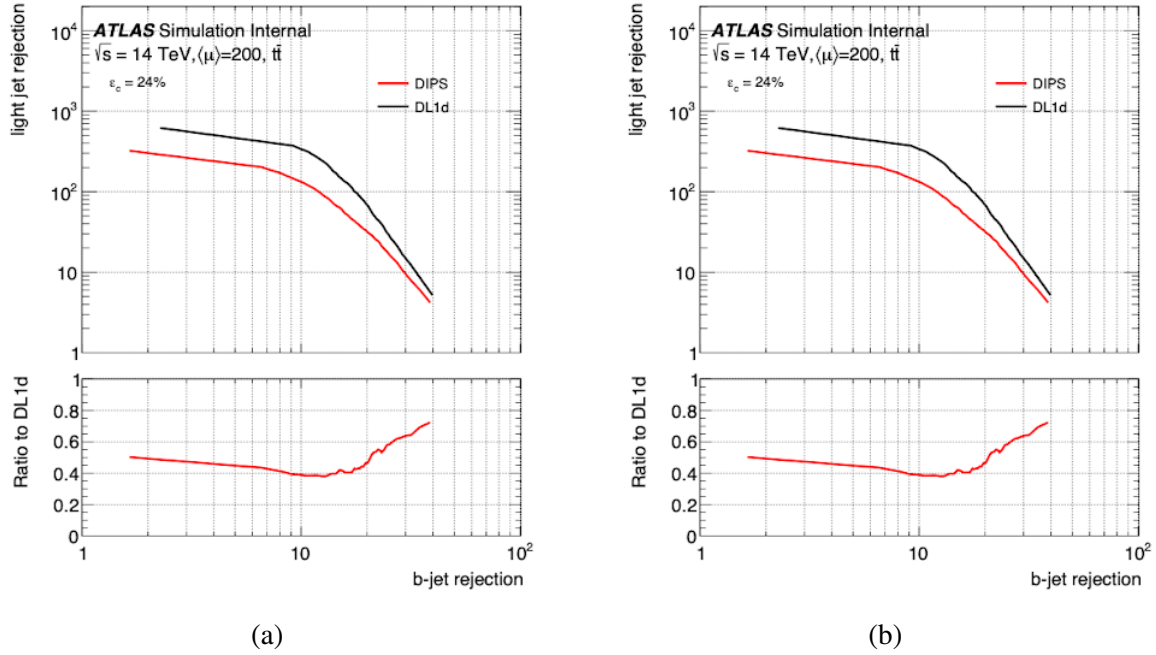
Figure 45: Light-jet rejection vs b-jet rejection (c-tagging). The left plot has a floating fraction of f_b=0.24 where as the right has f_b=0.45. The DL1d tagger outperforms DIPS as expected. The small HGTD sampled as listed in Table 13 was used.

include well resolved jets. The current state-of-the-art DL1d tagger was trained using 120 million jets simulated using Run 2 ATLAS geometry with well resolved jets between the eta range of $0 < |\eta| < 2.4$. The comparison of these two versions of DL1d can be seen in Figure 46. This figure also includes the old high-level tagger of MV2 that was validated on samples with the inclusive eta range of $0 < |\eta| < 4.0$. In this comparison ROC plot, it is seen that the DL1d for Run 2 outperforms the preliminary DL1d for Upgrade up to the 77% WP but starts to under perform at the 85% WP in the light-jet rejection. This result is highly promising knowing the robustness of the Run 2 version. If a more refined DL1d for Upgrade model is trained using an increase in statistics, it would be expected to outperform the Run 2 DL1d at the 77% WP.

Adding the newly available eta region of $2.4 < |\eta| < 4.0$ through the calculated power of the ITk detector has physicists excited about the new aspects that this adds to their analyses and possible innovations in the future. Though, in order to effectively probe this region, the available FTAG high-level taggers must be trained with robustness. Due to the decreasing resolution as particles become highly boosted in a more forward region of the detector, it is expected for particle tagging
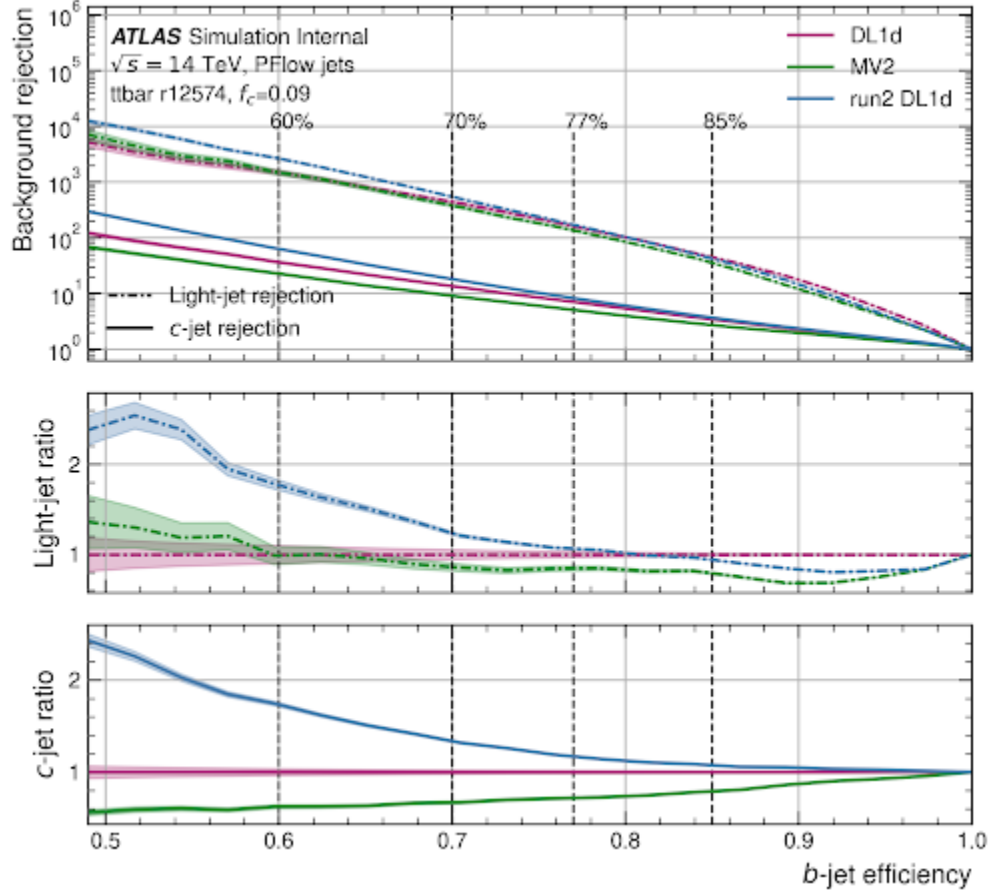
Figure 46: ROC curve comparing the performance of the DL1d for Run 4, the current version of DL1d used for Run 2 and the MV2 high-level tagger.

efficiency to drop. Thus, requiring newer innovative tools and techniques to be implemented with the hopes to obtain the maximum efficacy of this new phase space. The high-level tagger DL1d has been state-of-the-art through the end of Run 2 and into Run 3 of the ATLAS detector's campaign lifetimes. The question is how well will this tagger perform during Run 4 within the extended eta region. Figure 47 shows its current performance in increasing eta intervals of one. The performance in the highly boosted region between $3 < |\eta| < 4.0$ shows the poorest performance but this is to be expected. The Upgrade DL1d model was also validated on jets that are only contained in the eta region $0 < |\eta| < 2.5$ to have a proper comparison between the current implemented Run 2 DL1d tagger. The ratio plot on Figure 47 shows the performance between these two taggers. Similar performance is seen at the 77% WP with an increasing performance above this value. This result

is very promising since the Run 4 version of the DL1d tagger was trained on a magnitude less of statistics.

Overall, this first preliminary study of the high-level tagger DL1d using samples that simulate the Upgrade geometry of the ATLAS detector during Run 4 and the first implementation of the HL-LHC shows very promising results. There are currently active efforts to develop a new state-of-the-art high-level tagger using a graph neural network on tracking information called GN1. The architecture was briefly discussed in Section 4.6.4. Preliminary studies of this new tagger were done using the same samples that were used in this study of Run 4 for DL1d. The results of this study are discussed in Appendix APPENDIX A:. As seen from the results of this GN1 tagger, it outperforms the DL1d tagger. Due to this, a graph neural based tagger is planned to be considered the new industry used tagger after the upgrade in 2029. Therefore, further studies of the DL1d tagger using larger Upgrade samples is not planned for and the model trained in this dissertation will be used at the DL1d baseline model for Upgrade.
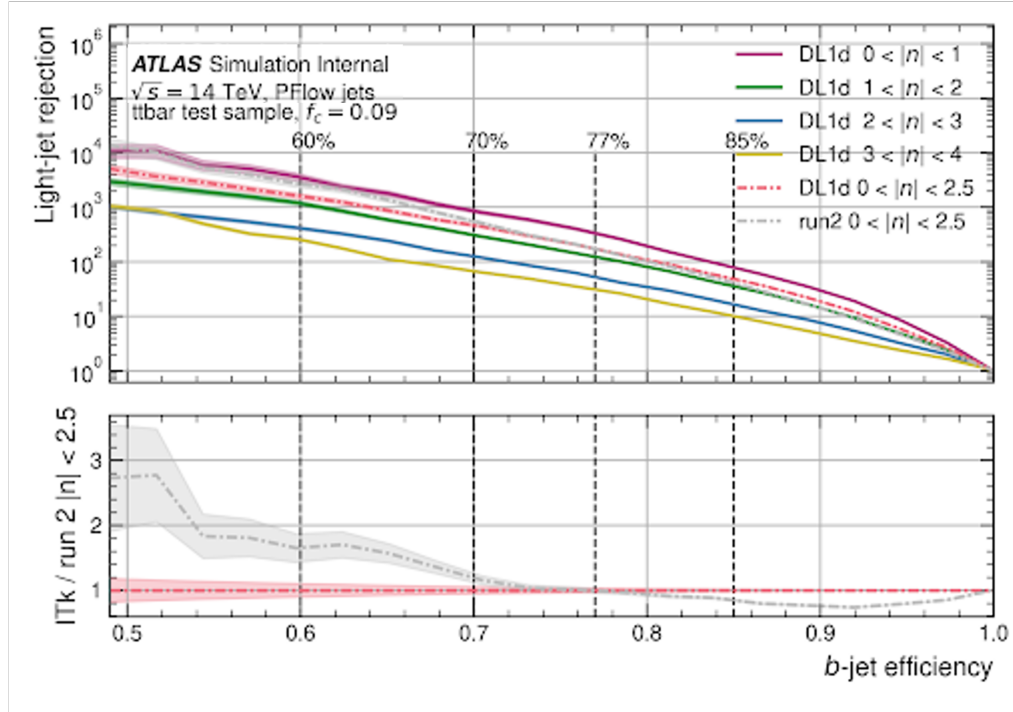
Figure 47: ROC curve showing the performance of the DL1d tagger for Upgrade in four eta intervals of one. The Run 2 DL1d is also plotted for comparison in the ratio plot on the bottom.

CHAPTER VI

SEARCH FOR NEW PHYSICS USING UNSUPERVISED MACHINE LEARNING FOR ANOMALY DETECTION

Anomaly detection methods are in their infant stages within the HEP community. This analysis was the first published general search using an unsupervised machine learning method within the ATLAS collaboration. This method searches for new physics in two-body invariant masses in events with a single isolated lepton using all data from Run 2 (140fb^{-1}) of $\sqrt{s} = 13$ TeV collision data recorded by the ATLAS detector. The invariant mass distributions are a phase space defined by an unsupervised deep-learning machine learning algorithm called the autoencoder (AE). The di-object invariant masses, m_{jX}, are constructed from a leading jet j and X is a second jet, a b-jet, lepton, or a photon. Similarly, m_{bX}, is looked at where b is a b-jet and X is a jet, b-jet, lepton or a photon, totaling in nine different invariant mass combinations. A search for di-object resonances was applied in the invariant mass range between 0.3 TeV to 6 TeV. The AE was trained using a data-driven technique, therefore did not rely on MC calculations. The invariant masses in the outlier phase space was set to a 95% confidence level upper limits on cross-section times branching ratios for the production of decays of resonances as predicted by several new physics scenarios.

6.1   Strategy

Due to the fact that this analysis was still paving the way to normalize anomaly detection methods and techniques, a few choices were fairly ambiguous and had to be thoroughly thought through in order to establish credibility. This method searching for signatures of BSM physics is model agnostic, meaning no theory for new physics influenced any of the model training and therefore any signatures detected is purely data-driven. The idea is to perform unsupervised training on a ML

model that will be able to detect events that differ from the "average" SM events, i.e. anomalous events. The input representation and training procedure should not bias the signatures and create artificial peaks. The strategy is this:

- Prepare the events into an input feature space that is general enough to cover a large range of possible BSM signatures while also providing an unbiased depth in data representation and

- Train an optimized ML model with metrics that can be used to define an anomaly score in order to show separation between "average" SM events and outliers.

- Define anomalous events using a chosen anomaly score and use a cut on this score to obtain an anomalous region of interest.

- Study these anomaly regions (ARs) for possible new physics signatures.

The last step suggests that new two-body states may be produced within these ARs and can be seen as excesses in these distributions, which can be found without using any MC simulations for background modeling.

6.2   Event Selection and Object Definitions

All the data used in this analysis is originated from the ATLAS detector during the Run 2 period between the years 2015 to 2018. The data was recorded during stable beam conditions while all relevant subdetectors were fully operational. The event candidates selected was done by either single-electron triggers or single-muon triggers which range in transverse momenta, transverse energy, and quality and isolation thresholds. These GRL files, datasets and used triggers can be found within the appendix section APPENDIX B:.

Muon and electron criteria are key to this analysis since they are the trigger for event selections. The Muon and electron selection criteria are summarized in Table 17 and Table 18, respectively.

104

Muon Selection	
Criteria	Value
Selection WP	medium
Isolation WP	PflowTight_FixedRad
Momentum Calibration	Sagitta Correction Not Used
p_T Cut	> 20 GeV
$|\eta|$ Cut	< 2.7
d_0 Significance Cut	< 3
z_0 Cut	< 0.5

Table 17: Muon selections for this analysis

Electron Selection	
Criteria	Value
Pseudorapidity Range	$(|\eta| < 1.27) \,\|\, (1.52 < |\eta| < 2.47)$
Energy Calibration	"es2018_R21_v0" (ESModel)
Transverse Momentum	$p_T > 20$ GeV
Track-to-Vertex Association	$|d_0^{BL}(\sigma)| < 5$
	$|\Delta z_0^{BL} sin\theta| < 0.5$ mm
Selection WP	Tight
Isolation WP	FCTight

Table 18: Electron selections for this analysis

6.2.1 Photon Selection and Reconstruction

Photon energy depositions are found within the ECAL which have a $p_T > 20$ GeV and have $|\eta| <$ 2.37. The transition region between the ECAL and the barrel, $1.37 < \|\eta\| < 1.52$, are excluded. There are two types of photons, one is a converted photon which are formed from ECAL clusters which are matched to a conversion vertex and then there are unconverted photons which are clusters that are not matched to any vertex. The energy depositions, i.e. shower shapes, undergo a strict criteria to pass selections that correspond to the *Tight* identification WP and *Tight* isolation. These photons are used in the invariant masses, $m_{j\gamma}$ and $m_{b\gamma}$.

6.2.2 Jet Definition and Selection

Jets are constructed using the anti-k_T algorithm with a distance parameter of R = 0.4. They are reconstructed from PFlow objects. In order to suppress pile-up, the JVT technique is used, requiring

at least 60% of the tracks momentum to be associated with the hard scatter. The JVT algorithm is applied to jets with p_T < 60 GeV and $|\eta|$ < 2.4. The final jet cut selection requires p_T > 20 GeV and $|\eta|$ < 2.4. Table 19 shows the definition of jets used in this analysis.

Jet Reconstruction Parameters	
Parameter	Value
algorithm	anti-k_T
R-parameter	0.4
input constituent	PFlow
Analysis Release Number	21.2.177
Selection Requirements	
Observable	Requirement
Jet Cleaning	LooseBad
BatMan Cleaning	No
p_T	> 20 GeV
$|\eta|$	< 2.47
JVT WP	Medium

Table 19: Jet definitions used in this analysis

6.2.3 B-jet Selection

The AntiKt4EMPFlow b-jets were selected from the 77% WP using the DL1r algorithm. The minimum jet p_T for pre-selection was 20 GeV. No optimizations of the WPs were made based on the BSM models used.

6.2.4 Final Selection

After all the pre-selection requirements and stated in the previous sections were made, a final selection is applied to define our signal region. This region requires at least one isolated lepton with p_T > 60 GeV and at least one jet with p_T > 30 GeV. The pseudorapidity requirement is $|\eta|$ < 2.4 for jets, $|\eta|$ < 2.5 for muons and $|\eta|$ < 2.47 for electrons. The FCTight and PFlowTight isolation WPs were used for electrons and muons respectively. This tight criteria imposed on leptons allows for consistent candidate definitions. The invariant masses, m_{jj}, m_{jb}, m_{bb}, are reconstructed using two jets from each event, either with a leading and sub-leading (b-)jets, or with a b-jet or anti-b-jet.

106

The minimum value of the invariant mass was chosen to be 400 GeV. This decision was made, in short, because the lepton cut of $p_T > 60$ GeV is expected to distort the mass distributions of m_{je} and $m_{j\mu}$ between the range of 200 and 400 GeV.

6.3   Monte Carlo Simulations

Even though this technique is considered a data-driven approach in order to not rely on MC simulations. Simulations were still used to describe the background hypothesis for a proper fit procedure. MC simulations were also produced for several benchmark BSM models that were used to show how well the trained AE could identify anomalous events that originate from new physics. A more in depth discussion on the types of MC samples used for the background hypothesis can be found in appendix APPENDIX C:.

6.3.1   Benchmark BSM Models

The following BSM models are used as benchmark signals to evaluate the efficacy of the unsupervised anomaly detection ML model. The Feynman diagrams are shown in Figure 48. The motivation of choosing these specific models are discussed in the following.

6.3.2   Sequential Standard Model

The Sequential Standard Model (SSM) is an extended gauge model [5] which proposes heavy gauge bosons which are commonly denoted as W$'$ and Z$'$. The emission of a W boson was considered from the s-channel production

$$qq \rightarrow \text{W}' \rightarrow \text{W Z}' \rightarrow (\text{l}\nu)(\text{qq}) \tag{6.1}$$

where Z$'$ is a new dijet resonance that is produced in association of the W boson [94]. The branching ratio W$'$ to Z$'$W was set to 50%, while 100% branching from $Z \rightarrow jj$ was set to increase the efficiency of MC production.

(a) Sequential Standard Model

(b) Simplified DM Model

(c) Radion Model

(d) Composite-lepton Model

(e) Charged Higgs Model
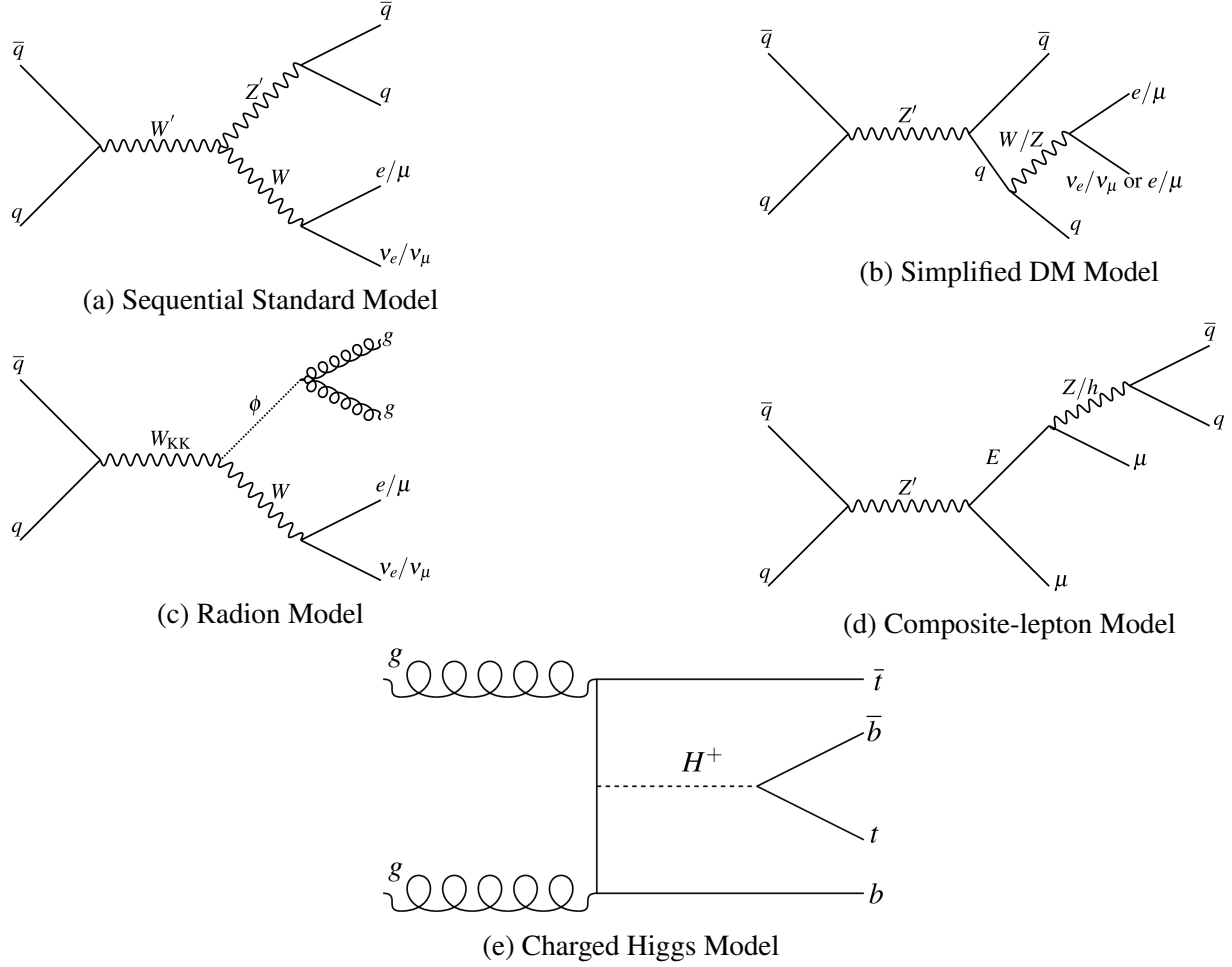
Figure 48: Feynman diagrams of the benchmark BSM models.

6.3.3 Simplified Dark Matter Model

The simplified dark matter (DM) models contain one or more stable, long-lived DM particle along with an unstable mediator particle that interacts between DM and the SM. The model used in this analysis consists of a single spin-1 mediator denoted as Z' created through a new U(1) gauge symmetry. The final states contain one or more leptons/

6.3.4 Kaluza-Klein Bosons Decaying to Radions

In order to solve the elector-weak hierarchy problem and flavor structure origin, some BSM theories predict warped higher dimensional compactifications with bulk SM. In the model used for this analysis, a Kaluza-Klein (KK) excitation gauge boson may decay into a particle called the radion

and a SM gauge boson [76, 51].

$$Wkk \rightarrow W + \varphi \rightarrow l\nu + gg \tag{6.2}$$

where Wkk denotes the KK boson and the φ is a radion decaying into two gluons.

6.3.5 Composite-Lepton Model

Composite resonances breaking lepton flavor universality predicts a Z' particle that decays into a composite lepton (E) and a SM lepton [19]. The composite lepton then decays into a lepton and a Higgs boson or Z boson.

$$Z' \rightarrow l + E; E \rightarrow e + Z/h; Z/h \rightarrow q\bar{q} \tag{6.3}$$

6.3.6 Charged Higgs Model

Many BSM models predict the existence of a charged Higgs boson. For this analysis, the simulated process assumes the charged Higgs boson is produced along with a top quark and a bottom quark [79]. It then decays itself into a top and bottom quark. In the very boosted regime, such as H^+ masses above 1 TeV, a b-jet and a jet originated from a top quark form almost back-to-back, which would be reconstructed via the m_{jj} distributions. For lower, non-boosted masses, the leading and non-leading jets lead to an approximate invariant mass of the H^+ and end in a rather broad resonance due to incomplete reconstruction of the decay products.

6.4   Event Input Representation

Now that the pre-selection has been chosen, the question is how to represent this as input for a machine learning model in order to maximize underlying correlations and pattern recognition. The overall goal of this analysis is to find anomalous dijet resonances, but the input must ensure that the model is not biased towards only kinematic anomalies. The approach that was taken for this analysis was to represent the input as a matrix that contains correlations between each object within

the event. This matrix is called the Rapidity Mass Matrix, or RMM. Figure 49 shows an example of the RMM with two object types, jets (j) and muons (μ). The maximum amount of objects is set to N. The position (1,1) contains the event's missing transverse energy, or MET, and is scaled by the center of mass energy ($1/\sqrt{s}$) where $\sqrt{s}$ is the center-of-mass energy. The diagonal cells contain the ratio $e_T(i_n) = E_T(i_1)/\sqrt{s}$, where $E_T(i_1)$ is the transverse energy of a leading object i (a jet or μ), and transverse energy imbalances

$$\delta e_T(i_n) = \frac{E_T(i_{n-1}) - E_T(i_n)}{E_T(i_{n-1}) + E_T(i_n)}, \quad n = 2, \dots, N, \tag{6.4}$$

for a given object type i. All objects are strictly order in transverse energy, i.e. $E_T(i_{n-1}) > E_T(i_n)$. The top row are the particle's transverse masses $M_T(i_n)$ for two-body decays, scaled by $1/\sqrt{s}$, i.e. $m_T(i_n) = M_T(i_n)/\sqrt{s}$. The upper-right quadrant shown in red are the non-diagonal values of $m(i_n, j_k) = M_{i,n,j,k}/\sqrt{s}$, where $M_{i,n,j,k}$ are two-particle invariant masses. The first column vector s $h_L(i_n) = C(\cosh(y) - 1)$, where y is the rapidity of a particle i_n, and C is a constant defined such that the average values of $h_L(i_n)$ can correspond to certain algorithms that may require values to have similar weights. The values in the bottom-left quadrant highlighted in green $h(i_n, j_k) = C(\cosh(\Delta y/2) - 1)$ are constructed from the rapidity differences $\Delta y = y_{i_k} - y_{j_n}$ between i and j.

Figure 49: Example of the Rapidity Mass Matrix using only two objects, jets (j) and muons (μ).

Applying this idea to this analysis expounds this example RMM into a larger representation. In the end, nine invariant masses were studied, therefore the RMM had to contain all the objects that

were used in combination for the invariant masses. The total objects per event varies event to event. In order to deal with this variable sizing, the events were "mapped" to a fixed data-structure and therefore implement zero-padding for missing data. The the standard topology of reconstructed objects was used while setting a maximum number for each object. In total, up to 10 jets, 10 b-jets, 5 electrons, 5 muons, 5 photons and MET were allowed (total of 36 objects). In order to reduce biasing the ML model on the di-object invariant masses of interest, the values that correspond these nine invariant masses are zero-padded for every event. This gives us a total amount of variables to feed the model of 1287 (36^2-9=1287). Figure 50 shows an example RMM with all indices allotted for objects to be filled and shown as yellow, whereas the zero-padded indices are shown in blue. The indices corresponding to the nine invariant masses of interest are removed to reduce bias. Appendix shows examples of single events converted to RMMs.
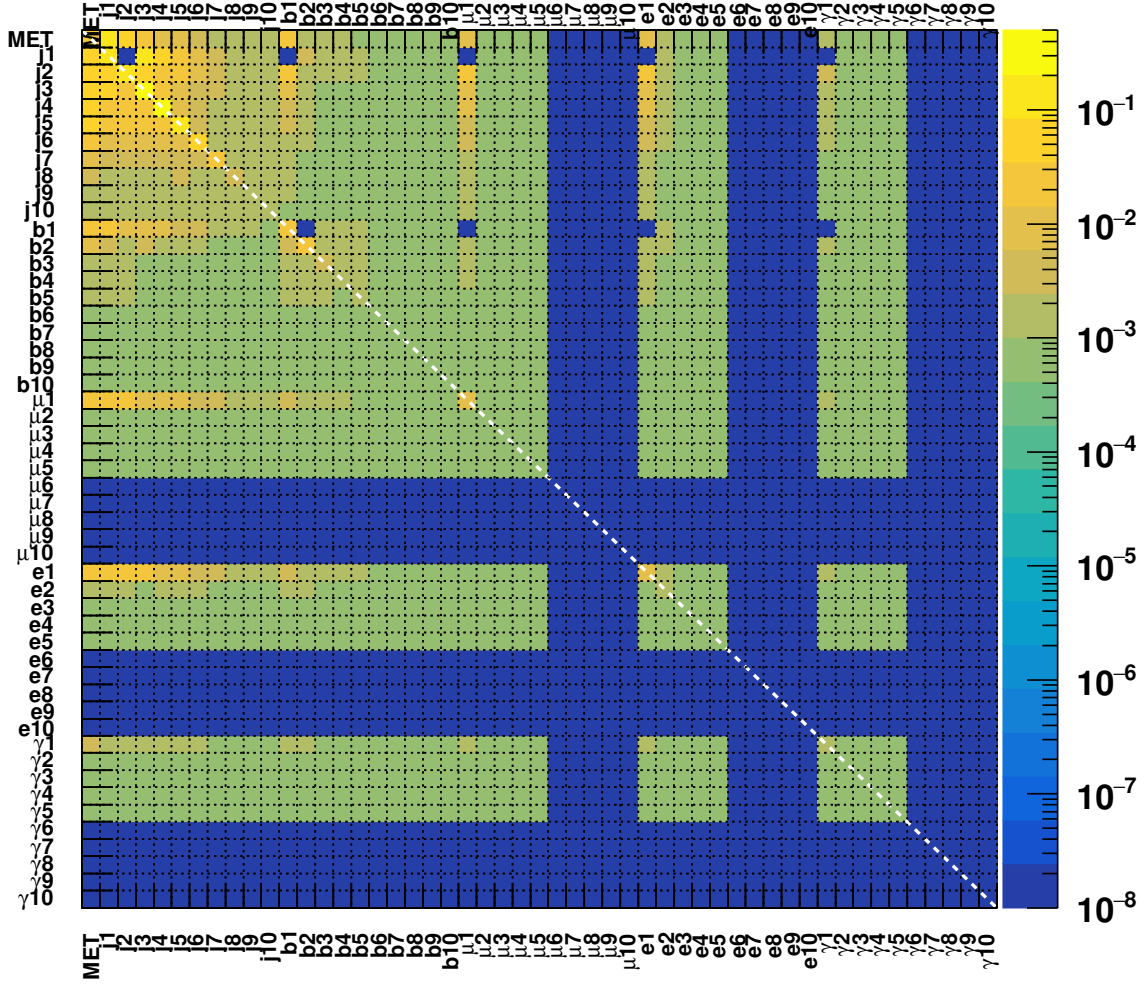
Figure 50: This RMM diagram shows the indices that allow values (yellow) and the zero-padded indices (blue). The nine invariant masses of interest are removed (blue). This diagram shows the average values of cells for 10000 events. The total of non-zero variables is 1287.

6.5 Autoencoder Training

The AE model was trained using `TensorFlow` [55] with a Keras backend [49]. The AE architecture is a deep-learning algorithm that is used for high-dimensionality reconstruction. The architecture is split into three parts. The first part is called the "encoder" which is the initial compression neurons which compresses the input into lower dimensionality. The second part of this architecture is referred to as the "latent layer" and is considered the bottleneck. This is a single layer of neurons that holds the compressed input. The final stage of this architecture is called the "decoder" which

decompresses the latent layer in order to reconstruct the original input. The neural-network of the decoder typically mirrors that of the encoder. The chosen amount of neurons should for these layers are optimized for the goal of anomaly detection.

The AEs purpose is to reconstruct its input. When the input gets compressed and then decompressed by the AE, there is a certain amount of data lost within the process. This value is determined by the loss function that the model is trained to minimize. The loss function used for this model is the mean squared error function (MSE) which can be seen in Eq. 6.5

$$Loss = \frac{1}{n} \sum_{i=1}^{n=1287} (x_i - \hat{x}_i)^2 \tag{6.5}$$

It is known for AEs that it's possible to have exactly zero loss when decompressing the input. This zero loss is not the goal of this architecture for the value of the loss is chosen to be the score in which indicates anomalous events. Therefore, the optimized architecture must have a range of loss values while maximizing the separation between SM events and BSM events. The architecture topology studies for this optimization can be found in Appendix APPENDIX E:. The resulting optimized architecture chosen is seen in Table 20. The total number of trainable weights is 2,863,087, the activation function used is the "leakyReLU" and the loss function is minimized by the Adam Optimizer. The reconstruction loss is used as the anomaly score.

```
--------------------------------------------------------------------
Layer (type)                  Output Shape              Param #

====================================================================

input_1 (InputLayer)          [(None, 1287)]            0

--------------------------------------------------------------------

dense (Dense)                 (None, 800)               1030400

--------------------------------------------------------------------

dense_1 (Dense)               (None, 400)               320400

--------------------------------------------------------------------

dense_2 (Dense)               (None, 200)               80200

--------------------------------------------------------------------

dense_3 (Dense)               (None, 400)               80400

--------------------------------------------------------------------

dense_4 (Dense)               (None, 800)               320800

--------------------------------------------------------------------

dense_5 (Dense)               (None, 1287)              1030887

====================================================================

Total params: 2,863,087

Trainable params: 2,863,087

Non-trainable params: 0
```

Table 20: Optimized Autoencoder architecture chosen for the anomaly detection analysis. More neurons may have optimized it further but was limited due to computational power.

Figure 51 shows a schematic of the nominal AE model with a sample input and output. 1% of ATLAS Run 2 data is randomly selected from different data-taking periods were used for training. 70% were used for training while 30% were used for validation. Early stopping was set to 30 epochs. MC samples nor labels were used for training, thus making this approach a data-driven unsupervised

learning. Figure 52 shows the training loss per epoch for both the training and validation sets.

$$Loss = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$$
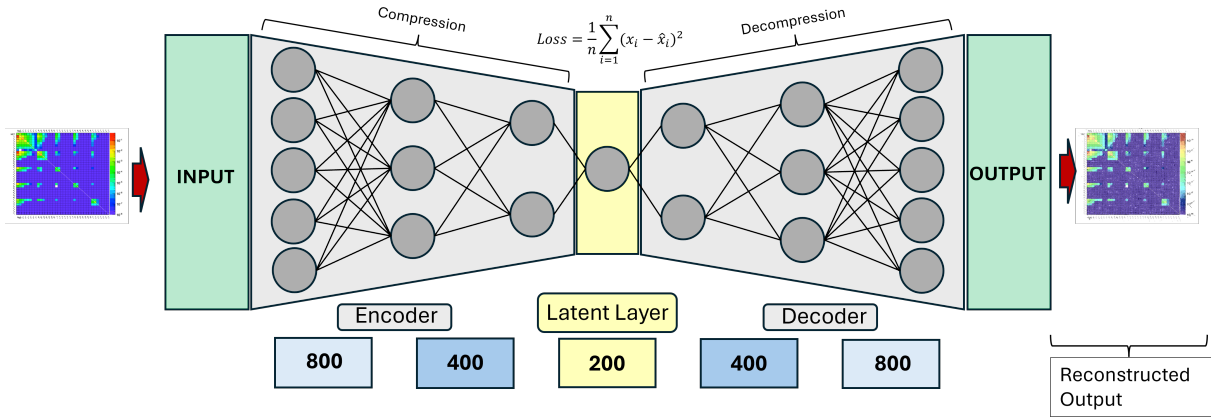
Figure 51: A schematic representation of the nominal AE model with an example input and its output. It's composed of three parts, the encoder which compress the data, the latent layer which acts as the bottleneck and the decoder which decompresses the data in order to recreate the original input. Due to this compression and decompression, data is loss via the loss function calculation. This loss is used as the anomaly score. When an event that is the model hasn't seen goes through, the data loss is higher and thus can be tagged as anomalous.
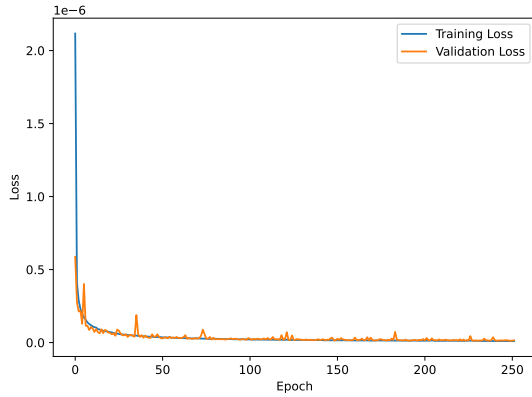
(a) Loss per epoch (linear)
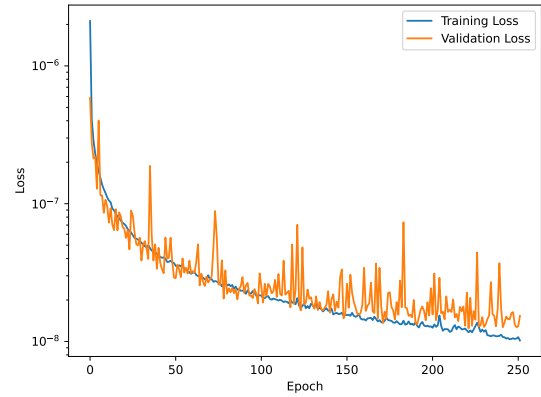
(b) Loss per epoch (log)

Figure 52: Training and validation loss as a function of epochs shown both linear and log y-scale.

Training ML models have dependencies on several factors such as random seed values used for AE initialization, the processing computer architecture, training/validation dataset splitting, etc. In order to accommodate for these factors, 50 separate trainings were conducted with different random seed values. Figure 53 shows these 50 trainings and the validation value in which the AE was stopped training. The median value of this validation loss is ln(`loss`) is -10.554. The trained
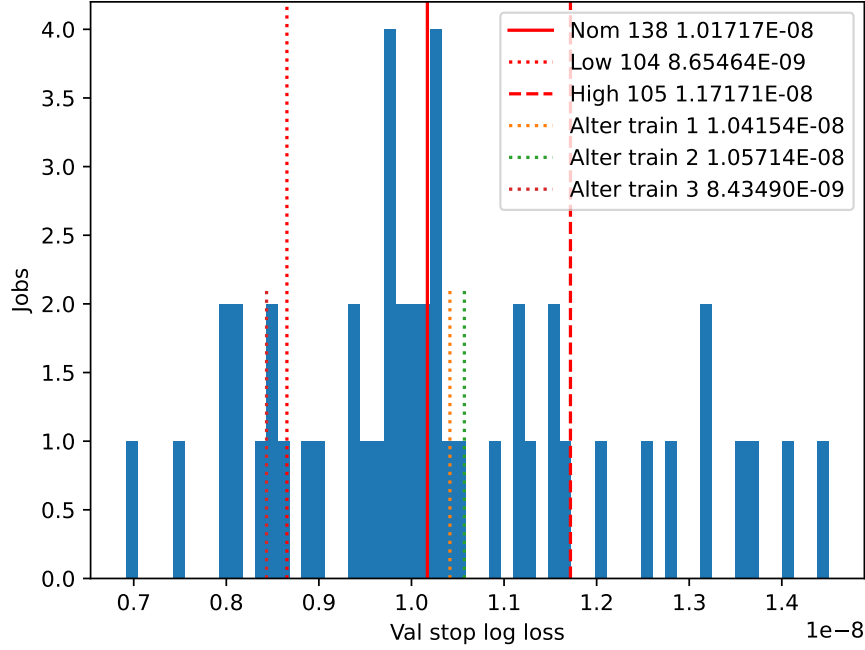
(a) Using 1% of data for training

Figure 53: Validation stop loss values for 50 trainings with different seeds. Seed value and corresponding loss value is shown in the legend. "Alter training" shows different randomly selected training data from full Run 2 ATLAS data. (ROC curves can be found in Appendix APPENDIX F:) Consistent loss values were achieved among different training sets.

model that stopped at this median value then became the nominal model used for the rest of the analysis. The two ±RMS models are used for systematic studies in Appendix APPENDIX F:.

The trained nominal AE model was used to process MC events and BSM events and is shown in Figure 54. The ln(loss) value is taken to show anomalous separation. 10% of randomly chosen data events from Run 2 were used to check the loss distribution to ensure that there is good separation between data and BSM. It can be seen that in both (a) and (b) in Figure 54 that there is good separation between SM and BSM events. The year dependence is also checked since all of Run 2 is taken over several years. Figure 55 shows that there is no shape dependence on data-taking years.
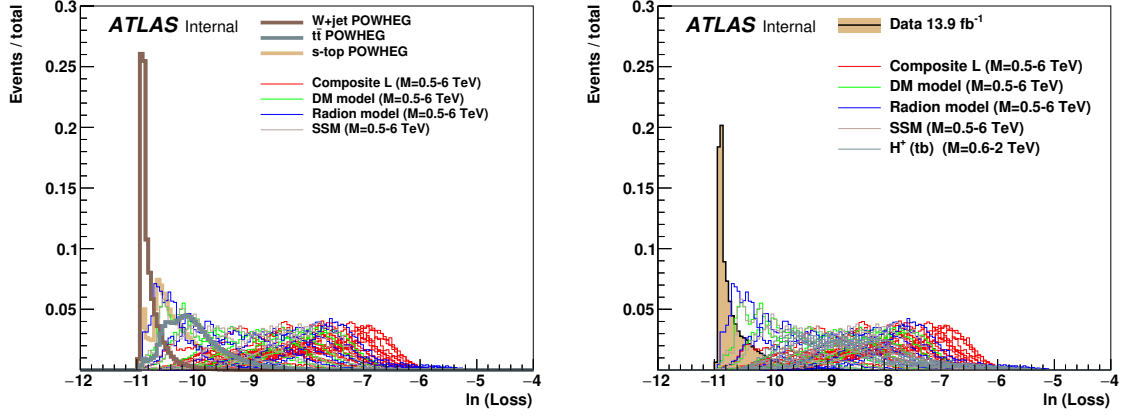
(a) MC for autoencoder trained with 1% of data   (b) Data for autoencoder trained with 1% of data

Figure 54: Distributions of the loss values for the AE trained using 1% of data. (a) The loss distribution for SM and BSM models; (b) The loss distribution for 10% and BSM models. The BSM models had 20,000 generated events for each mass point in the range 0.5 − 6 TeV. The larger the mass of the resonance, the further away the line is from the data distribution. All the distributions are normalized to the unit area.
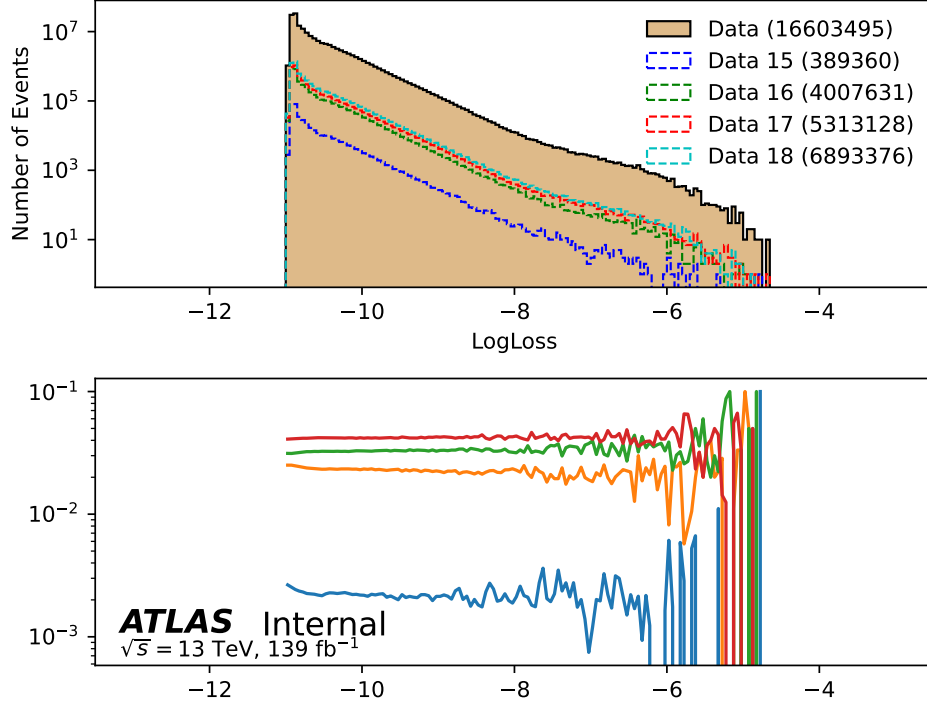
Figure 55: Distributions of the loss values for 10% Run 2 data, scaled by 10 to simulate the actual expected distribution. Distributions from each data taking year are shown as well (without multiplying by 10). To quantify the difference, the AE value between individual-year and all-year shapes are computed; they are found to be 'Data 15': 0.485, 'Data 16': 0.491, 'Data 17': 0.505, 'Data 18': 0.502. Ratio pad shows the per year shape over full data (note that colors may not be precisely matched). They are all close to 0.5, suggesting that different data-taking conditions (e.g. pile-up) do not have a large impact on the method.

6.5.1 Determination of Anomaly Region

The definition of an anomaly region (AR) is ambiguous and can be determined depending on the logical needs of the analysis. Within this analysis, the decision was based on the example BSM models using their theoretical cross-sections times their branching ration (after their acceptances time efficiency corrections). The $\sigma \times BR \times Acc \times Eff$ were calculated for these models and can be seen in Table 21. The range of these values are within a few hundredths to a few pico-barn. Three ARs were defined to ensure sensitivity to a wide range of BSM models. These three regions are event count bases ranging from 10 pb to 0.1 pb.

- **10 pb region:** This 10 pb AR corresponds to 10 pb ×140fb⁻¹=1.4M events. This region has

118

the largest acceptance and is a reasonable choice for many models with leptons in the final state.

- **1 pb region:** This 1 pb AR corresponds to 0.14M events of data. This region is less sensitive due to the decrease in statistics but it was expected that the same fit function as the 10 pb AR could be used. This assumption is justified since this region has less events and therefore an equation with the same amount or less parameters as the first region should work.

- **0.1 pb region:** This is the third and last AR definition and only corresponds to 14K events. This region was included since it corresponds to the upper bound on published experimental limits for dijet mass (including masses with associated lepton or photons). The same fit function will be used for this region as was used in the previous regions.

Since these ARs were defined using a count-based assumption, it was possible to define the anomaly score cut numerically. The values were found scanning the amount of events bin by bin on the log(loss) anomaly score distribution. The cut values for the 10 pb, 1 pb and the 0.1 pb ARs are -9.10, -8.00 and -6.5 respectively and be seen within Table 22.

Theoretical Cross Sections times Branching ratios for BSM Models			
$\sigma \times$ **BR** (pb)	$\sigma \times$**BR**$\times Acc \times Eff$ (pb)	Model	Reference
10	2.5	Sequential Standard Model	ATLAS [38]
1.5	0.2	Charged Higgs (hMSSM, **tan(β) = 1**)	ATLAS [38]
6	0.3	Charged Higgs (hMSSM, **tan(β) = 0.5**)	ATLAS [38]
0.5	0.14	Simplified dark matter model	ATLAS [38]
8 (extr)	1 (extr)	Composite lepton model (E=250 GeV)	ATLAS [40]
1.5	0.4 (extr)	Composite lepton model (E=500 GeV)	ATLAS [40]
0.12	0.05	Technicolor model	ATLAS [38]
4.5 (extr)	0.7 (extr)	Radion model	ATLAS [40]

Table 21: Theoretical cross-sections times branching ratios and after $Acc \times Eff$ corrections of multiple BSM models near the 400 GeV mass scale that include a single isolated lepton in the final state. The proposed ARs of 10 pb, 1 pb and 0.1 pb covers the cross-sections of most of these models.

Cut values	Number of events				Cut values	Number of events
−9.00	1146820.0		Cut values	Number of events	**−6.40**	11260.0
−9.04	1286950.0		**−7.95**	132090.0	**−6.45**	12450.0
−9.10	**1382880.0**		**−8.0**	**143730.0**	**−6.50**	**13520.0**
−9.12	1626590.0		**−8.02**	157050.0	**−6.55**	14770.0
−9.15	1828930.0		**−8.10**	171040.0	**−6.60**	16020.0
−9.20	1828930.0				**−6.65**	17600.0

Table 22: Number of events after the anomaly score cut for each AR. The 10 pb BSM region is defined by the logarithm of the loss function > -9.10, the 1 pb is defined by the logarithm loss > -8.00, and likewise for the 0.1 pb logarithm loss > -6.50

Multiple mass hypotheses of each example BSM model was observed using these defined AR regions in order to test the $S/\sqrt{B}$ after the anomaly score cut. Figure 56 shows comparisons for each five BSM model previously described. These mass hypotheses of $Z'/W'/H^+$ range from $0.5 - 6$ TeV. The larger the mass resonance (closer to 6 TeV), the larger the loss from the AE is expected. The number of events for each model were estimated using their cross-sections and the luminosity. Figure 57 shows the integrated S/B and $S/\sqrt{B}$ for all the BSM signals stacked. The 10 pb cut happens to be near the most optimal $S/\sqrt{B}$ meaning the signal in the calculation is dominated by lower mass resonances. It is worthy to note that this optimal cut would adjust towards the right from its current position if this calculation only includes higher mass resonances. The fact that the optimal cut is found to be at 10 pb for an inclusive mass resonance range indicates this method with the addition of the other two defined ARs is a robust strategy.
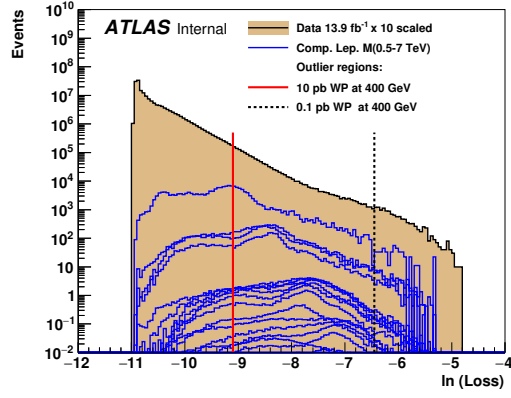
(a) Composite Leptons
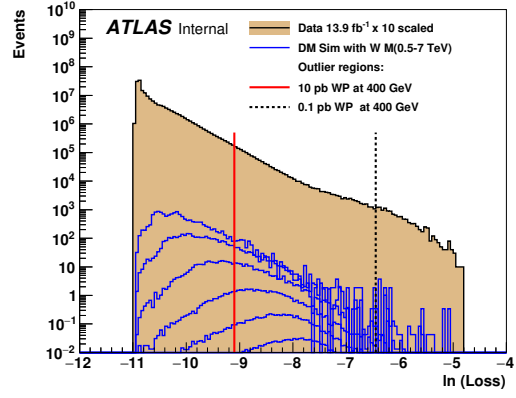
(b) Dark Matter
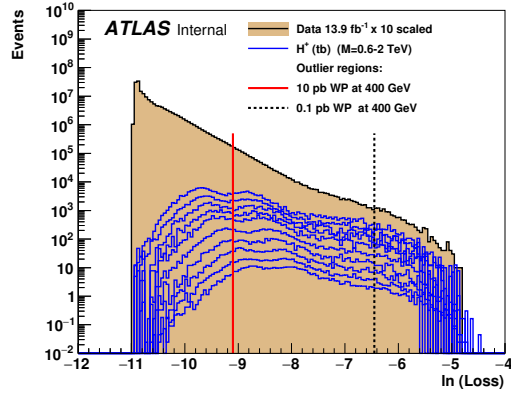
(c) Charged Higgs
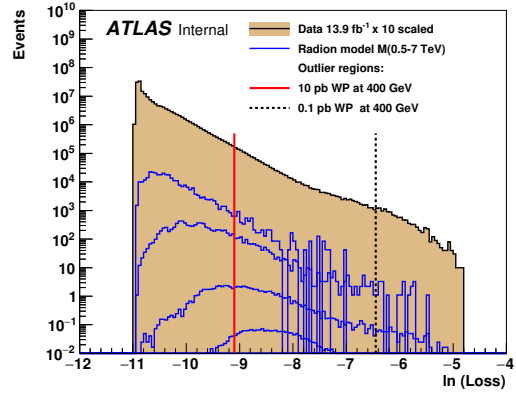
(d) Radion
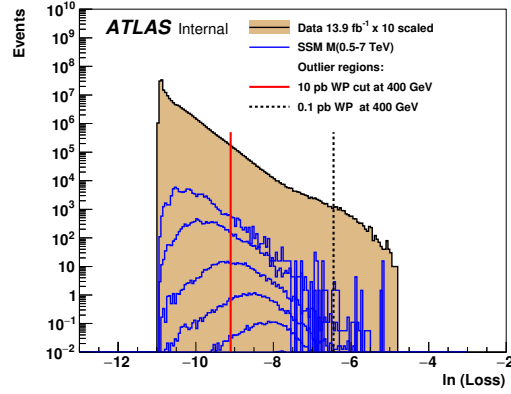
(e) Sequential SM

Figure 56: Distributions of the loss values using the trained nominal AE with different BSM models. The mass resonance range between 0.5 - 6 TeV with higher loss values for larger mass resonances. The data distribution uses 10% of Run 2 data that is scaled by 10 to show the expected shape of the full Run 2 dataset. Two vertical lines shows the max and min AR regions (10 pb and 0.1 pb)
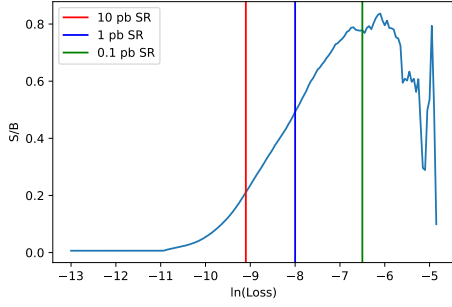
(a) S/B for stacked BSM models
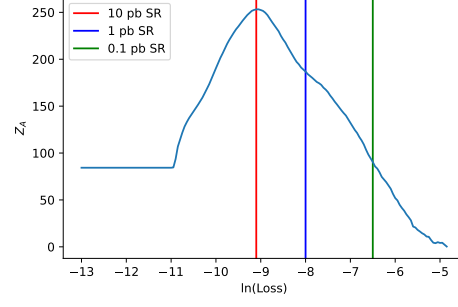
(b) S/$\sqrt{B}$ for stacked BSM models

Figure 57: Integrated S/B and S/$\sqrt{B}$ scans for the stack BSM models show in Figure 56. S is calculated as the integral of the BSM signal from a given x value to +infinity; B is calculated as the integral of the data at a give value x to +infinity. The colored vertical lines show the positions of the AR cuts.

6.6   Analysis of Anomaly Regions

The objects of interest are nine di-object invariant masses. These masses are:

1. Invariant mass of 2-jets (m_{jj})

2. Invariant mass of a jet and a b-jet (m_{jb})

3. Invariant mass of two b-jets (m_{bb})

4. Invariant mass of a jet and an electron (m_{je})

5. Invariant mass of a jet and muon ($m_{j\mu}$)

6. Invariant mass of a jet and photon ($m_{j\gamma}$)

7. Invariant mass of a b-jet and an electron (m_{be})

8. Invariant mass of a b-jet and a muon ($m_{j\mu}$)

9. Invariant mass of a b-jet and a photon ($m_{j\gamma}$)

The approach here is to use MC simulations of $t\bar{t}$ and W+jets (samples listed in Appendix AP-PENDIX C:) combined with a data-driven multi-jet for SM background estimations. With this

background a fit function is applied to establish the estimated shape of this background. The function is then tested further using a variety of statistical tests and then finally cross-checked with a 10% data sample (separate events from training) in a step called a stage-1 unblinding before being fully unblinded.

6.6.1 Background Modeling

The bin widths of the invariant masses are chosen to match the JES of the ATLAS detector. The bin widths are approximately equal to the theorized model width at a given mass with increasing bin size that widen with increasing invariant mass from 13 GeV to 120 GeV. The fit procedure and function in this analysis follows the procedure outlined in [42]. This fit hypothesis (Eq. 6.6) is used to establish the background shape.

$$f(x) = p_1(1-x)^{p_2} x^{p_3 + p_4 \ln x + p_5 \ln^2 x}, \tag{6.6}$$

where $x \equiv m_{jj}/\sqrt{s}$ and the p_i are five free parameters to be estimated. Having fit function with up to five parameters allows for more flexibility at low m_{jj}. This fit function from herein is referred to as "p5". In addition, in addition "p4" ($p_5 = 0$) and "p6" functions (after multiplying the function by $x^{p_6 \ln^3 x}$) are studied. Using a function with more parameters, such as the p6 can be dangerous as high parameter functions are more likely to absorb possible signal. Therefore, statistical tests are utilized to ensure the optimized amount of parameters are found while also testing the function on each invariant mass distribution since it's possible the fit could fail on a distribution with low statistics.

An alternative fit function is proposed to evaluate the systematic uncertainties related to the function to fit the background. The chosen alternative function can be seen in Eq. 6.7 as it resembles the used p5 fit function that is later used to estimate the background. This function has shown to give an alternative description of the tail in the m_{jj} distribution, compared to the p5 function [41].

$$f(x)^{alt} = p_1(1-x)^{p_2} x^{p_3 + p_4 \ln x + p_5/\sqrt{x}}, \tag{6.7}$$

124

To fit the binned histograms, the Minuit/ROOT program was used. histograms with event yields were divided by the bin width, this was fed as inputs into the Minuit/ROOT program to obtain fit parameters. To ensure the most optimal minimization of the fit function, the following procedure was used. First, the fit parameters are initialized randomly and the fit is performed. The fit criteria requires a χ^2/ndf <1.4 to be considered as the correct parameter space. If the fit doesn't converge to this criteria, a new fit trial begins starting from the parameters as the previous fit. 100 of these trials are then performed, if a fit cannot be found with a χ^2/ndf <1.4, the parameters are then randomly initialized and the procedure starts over. This iterative fitting procedure is then repeated a maximum of 100 times to obtain correct parameter space. If the criteria of χ^2/ndf <1.4 is never obtained, then the claim is that this fit function is unable to describe the background shape. /par Additional tests are performed once reasonable fitting parameters are found via this method. To ensure the function isn't "too flexible", one of these tests are spurious signal tests which are conducted to help verify the chosen form of the function. This procedure is non-trivial due to the fact that the chosen number of parameters must work for all nine invariant mass background shapes. If there are not enough parameters, the shapes won't converge into a viable fit, whereas if there are too many variables, it's possible that an hints of resonances may be hidden within the fit.

6.6.2   Statistical Tests

Several statistical tests are ran to ensure the quality of fit. These tests are ran on the normality of distributions of scaled residuals, $S_i = \frac{\mathrm{yields-pred}}{\Delta \mathrm{yields}}$, or commonly referred as "pulls". These pull values are then tested to verify that they follow a normal distribution with $\sigma{=}1$ and mean of 0. Seven tests are ran in all. These tests and their criteria are:

1. χ^2/ndf <1.4

2. Kolmogorov-Smirnov (KS) test has a probability > 0.3

3. Gaussian fit of the pulls has a mean of 0 within the 1σ statistical uncertainty

4. Gaussian fit of the pulls has a standard deviation with the value of 1 (within the 1σ uncertainty)

5. Skewness value is 0 within the 1σ uncertainty

6. Kurtosis value is 0 within the 1σ uncertainty

7. Shapiro-Wilk's test shows the probability above 0.7

First, as Gaussian fit to the pulls S_i is performed to make sure that the peak and the width of the fitted Gaussian has a mean of 0 with a width of 1. A χ^2 test is ran to verify that the Gaussian fit of the S_i is acceptable. The Shapiro-Wilk's test is used to detect all departures from normality without using the Gaussian fit. The Kolmogorov-Smirnov test is used to determine if the pull data is normally distributed. In addition, *skewness* is checked that measures the relative size of both tails along with the *kurtosis* test that measures whether if the distribution is heavy-tailed relative to the normal distribution. The skewness and Kurtosis tests are equal to zero for an normal distribution, so if the pull's tests are not close to zero then the test is considered failed.

To determine whether the number of chosen parameters for the function is optimized, the F-test is utilized. This test verifies if the default function against alternative, more complex options [16]. If no significant improvements are seen for the MC control regions, the function with lower complexity is prioritized since lower parameter functions lead to more stable fits and less likely to lead to spurious signals. All tests described here are performed on the SM MC plus the data-driven multi-jet background. In addition, the F-test is performed on real data in the 1 pb and the 0.1 pb to reduce parameters.

Once a fit function is tested and has proved to be viable for background estimation, a BumpHunter (BH) statistical test is performed to check if there are any significant deviations from the background, i.e. possible BSM resonances. This test is calculated by looking at the discrepancy in windows of varying width for all data bins and is sensitive to whether discrepancies in neighboring bins have sign outcome. This test also reports a final p-value as all possible locations for an excess is considered, choosing the lowest value. The frequentist approach that is used in this analysis

involves calculating this BH statistic for many pseudo-experiments, having the lowest p-value out-come originate from different locations. The final p-value is formed from the union of the smallest p-values calculated from these pseudo-experiments.

When applying the BH test to our ARs, the procedure looks for each N possible intervals of m_{jj}, a local p-value is calculated from a unique hypothesis test statistic. BH then combines each of the N hypotheses for form a new hypothesis test, calculating the minimum p-value amongst all the tests. These are then combined into a global p-value and transformed into a significance assuming that the bin-by-bin distribution follows a Poisson distribution. Pseudo-experiments as described above are then used to determine the most significant local excess and a global significance. This test is used after a reasonable fit function is found to describe the background estimation in order to find any deviances after unblinding.

6.6.3 Background Fit Studies

The background estimation in these studies are a combination of $t\bar{t}$, W+jets and the LE-CR. These figures show all nine invariant mass distributions in the 10 pb AR (fit studies looking at the other ARs, 1 pb and 0.1 pb, can be found in Appendices APPENDIX H: and APPENDIX I:). Three fit functions were tested, a p4, p5 and a p6. The statistical tests and discussed in the previous section for the pull values are shown in Table 23 for the 10 pb AR (Table 26 and Table 27 in Appendix for the other two ARs). These studies show that the p5 performs the best. The three fit functions tested are:

- $p4$: $f(x) = p_1(1 - x)^{p_2} x^{p_3 + p_4 \ln x}$;

- $p5$: $f(x) = p_1(1 - x)^{p_2} x^{p_3 + p_4 \ln x + p_5 \ln^2 x}$;

- $p6$: $f(x) = p_1(1 - x)^{p_2} x^{p_3 + p_4 \ln x + p_5 \ln^2 x + p_6 \ln^3 x}$.

To construct the MC+LE-CR, the distribution of the LE-CR was scaled such that it fills the gap between data (before AR cut) and the MC predictions for $t\bar{t}$ and W+jets. The data used was 10% of Run 2 and then scaled by 10 to show the full dataset shape as seen in Figure 58. The scaling

factor applied to the LE-CR is (*data-MC*)/*a*, where *a* is the value of the integral of the LE-CR. After applying the 10 pb AR cut to the combination of the MC+LE-CR, the fitting procedure for all the tested functions was performed on each invariant mass. The results of the statistical tests for these three functions on the 10 pb AR can be seen in Table 23. The studies on the 1 pb and 0.1 pb can be found in AppendixAPPENDIX H:-APPENDIX I:.
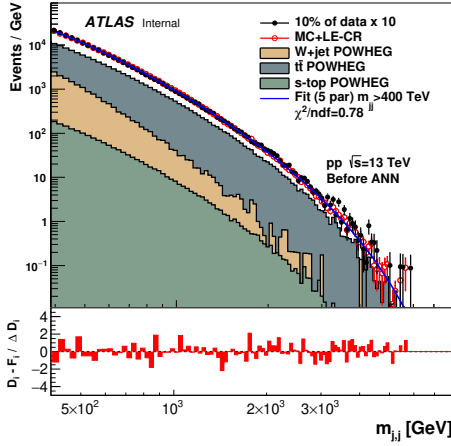
(a) Before 10 pb AR cut
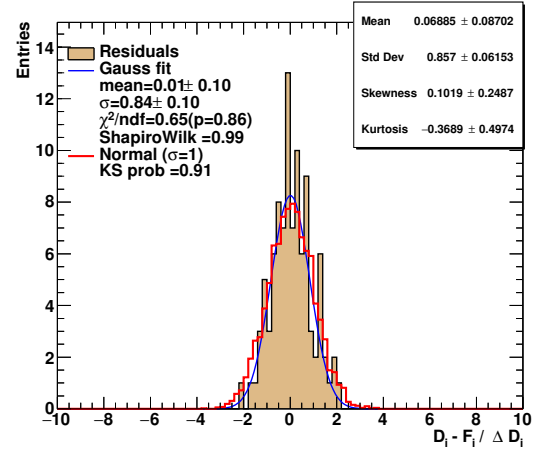
(b) Pulls before 10 pb AR cut

Figure 58: The distribution of m_{jj} before the 10 pb AR cut. The red dots show the MC+LE-CR data together with the p5 fit.

- Figure 59 10pb m_{jj}
- Figure 60 10pb m_{jb}
- Figure 61 10pb m_{bb}
- Figure 62 10pb m_{je}
- Figure 63 10pb $m_{j\mu}$
- Figure 64 10pb $m_{j\gamma}$
- Figure 65 10pb m_{be}
- Figure 63 10pb $m_{b\mu}$
- Figure 67 10pb $m_{b\gamma}$

128

Mass	region	p	Fit χ^2	Pull μ	$\Delta\mu$	σ	$\Delta\sigma$	Gaus χ^2	KS	Shapiro
m_{jj}	10 pb	4	1.350225	0.188249	0.118766	1.000243	0.118738	0.668667	0.838354	0.967003
m_{jb}	10 pb	4	1.654758	0.288073	0.174609	1.184496	0.190722	1.266895	0.204681	0.947339
m_{bb}	10 pb	4	1.339481	0.259308	0.184551	1.060607	0.198495	1.115230	0.030365	0.931438
m_{je}	10 pb	4	1.882063	0.181406	0.190882	1.402343	0.218868	0.983200	0.122363	0.938806
$m_{j\mu}$	10 pb	4	0.954052	0.267345	0.109163	0.742488	0.134621	0.777527	0.349305	0.945133
$m_{j\gamma}$	10 pb	4	0.773106	0.002398	0.130570	1.014041	0.135647	0.469318	0.798588	0.983349
m_{be}	10 pb	4	0.972373	0.545587	0.291598	1.414412	0.274838	0.768124	0.610562	0.970631
$m_{b\mu}$	10 pb	4	1.494746	0.706422	0.423572	1.646267	0.487647	0.663357	0.103588	0.967773
$m_{b\gamma}$	10 pb	4	0.824516	0.218063	0.156917	0.965077	0.163884	0.914854	0.200394	0.962123
m_{jj}	10 pb	5	1.009592	0.145683	0.122463	0.986835	0.109409	0.841748	0.915870	0.978346
m_{jb}	10 pb	5	1.195392	0.167570	0.169755	1.194622	0.210126	1.141865	0.449203	0.991612
m_{bb}	10 pb	5	0.903178	0.314245	0.105781	0.800464	0.094230	0.515502	0.302650	0.954691
m_{je}	10 pb	5	1.226098	0.185050	0.126940	0.989509	0.112932	1.166452	0.455442	0.975749
$m_{j\mu}$	10 pb	5	0.969384	0.402821	0.135216	0.923938	0.133419	0.808335	0.061480	0.936338
$m_{j\gamma}$	10 pb	5	0.781334	0.028170	0.130375	1.004669	0.138376	0.590685	0.630669	0.983825
m_{be}	10 pb	5	0.795931	0.331966	0.188702	1.092156	0.177057	0.671857	0.931749	0.973960
$m_{b\mu}$	10 pb	5	1.115614	0.594871	0.411939	1.518365	0.440538	0.744476	0.365553	0.976348
$m_{b\gamma}$	10 pb	5	0.696946	0.063387	0.120639	0.860113	0.104889	0.402123	0.926549	0.987653
m_{jj}	10 pb	6	0.954477	0.166474	0.114638	0.936988	0.106383	0.956339	0.974623	0.985057
m_{jb}	10 pb	6	1.397667	0.326289	0.174342	1.299782	0.192943	0.643885	0.614734	0.989053
m_{bb}	10 pb	6	0.950991	0.385546	0.109464	0.771136	0.096318	0.869510	0.236613	0.958079
m_{je}	10 pb	6	1.128457	-0.030160	0.149119	1.124779	0.143738	0.824662	0.615372	0.983419
$m_{j\mu}$	10 pb	6	0.978938	0.399265	0.135440	0.926759	0.147424	0.757382	0.061480	0.942393
$m_{j\gamma}$	10 pb	6	0.796370	0.034736	0.125878	1.017938	0.116741	0.414805	0.798588	0.986012
m_{be}	10 pb	6	0.833591	0.058978	0.135988	1.002012	0.120313	0.532795	0.882231	0.976117
$m_{b\mu}$	10 pb	6	1.146550	0.162036	0.209187	1.297957	0.215131	0.492555	0.230690	0.979333
$m_{b\gamma}$	10 pb	6	0.757415	0.081991	0.131751	0.905731	0.117748	0.488251	0.867562	0.983667

Table 23: Statistical quantities for SM MC+LE-CR fit for the 10 pb AR

(a) m$_{jj}$ using MC+LE-CR, p4

(b) pulls of m$_{jj}$ in p4

(c) m$_{jj}$ using MC+LE-CR, p5

(d) pulls of m$_{jj}$ in p5

(e) m$_{jj}$ using MC+LE-CR, p6

(f) pulls of m$_{jj}$ in p6

Figure 59: The m$_{jj}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.
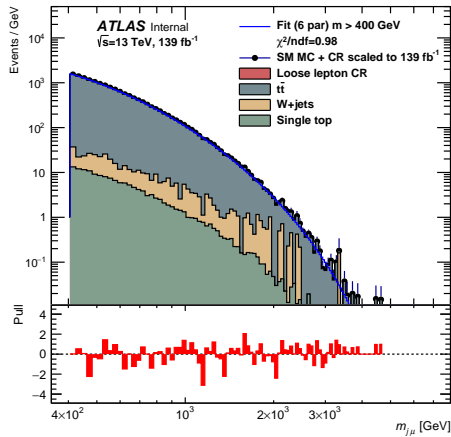
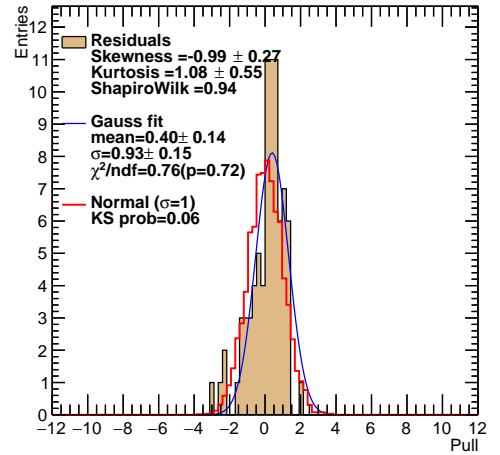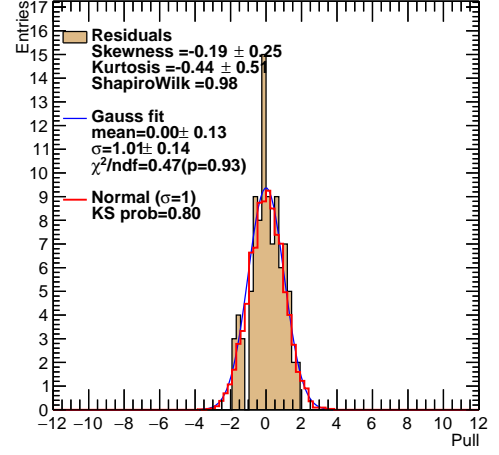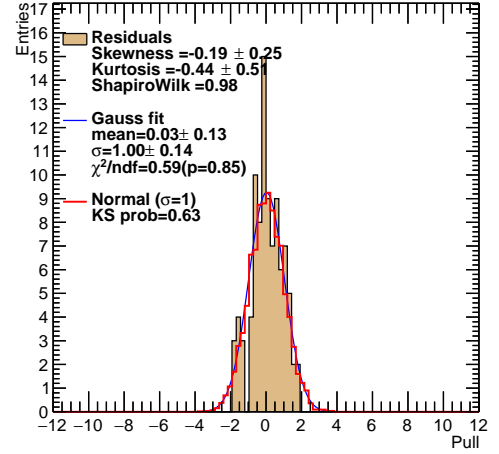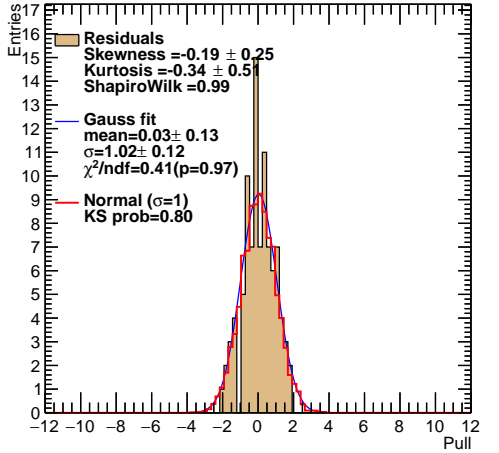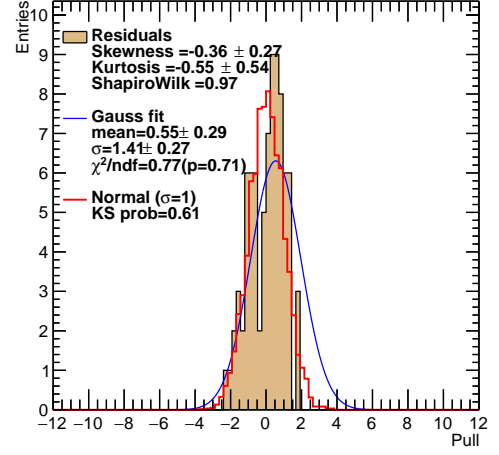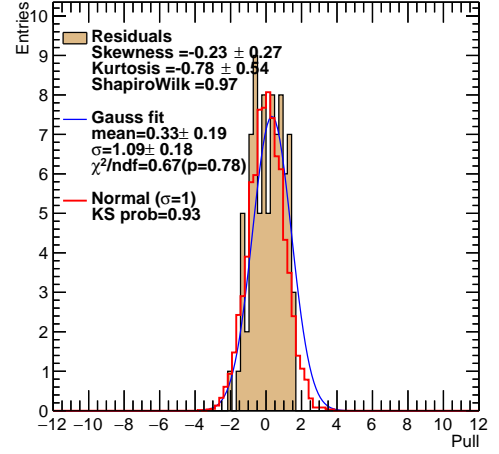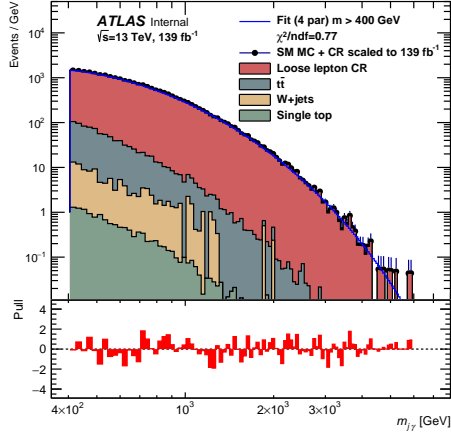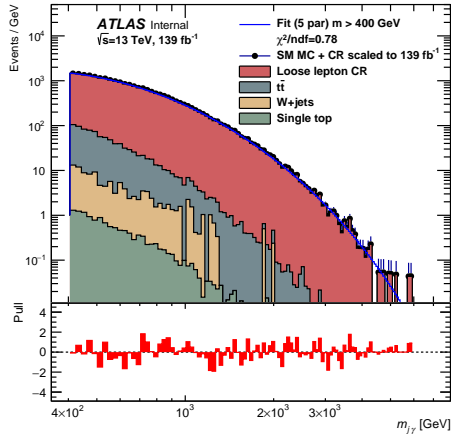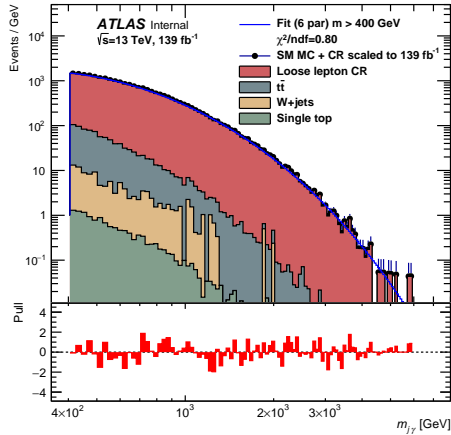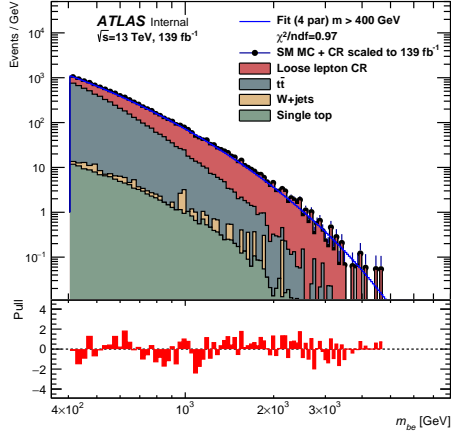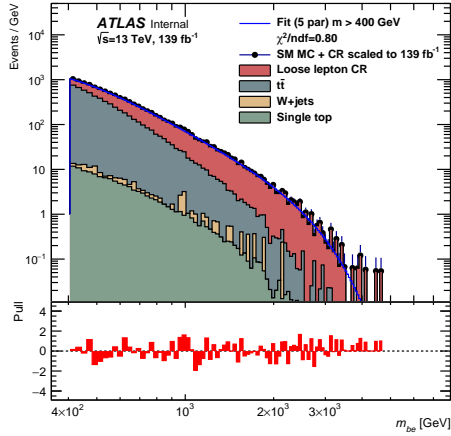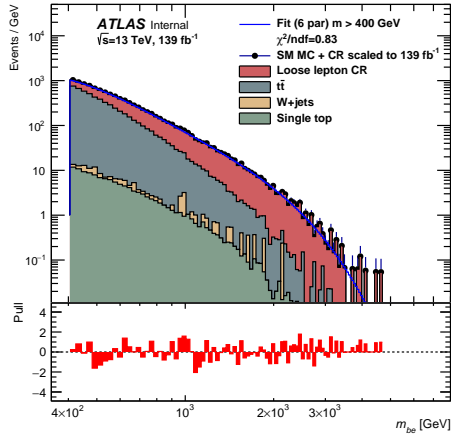(a) m$_{jb}$ using MC+LE-CR, p4
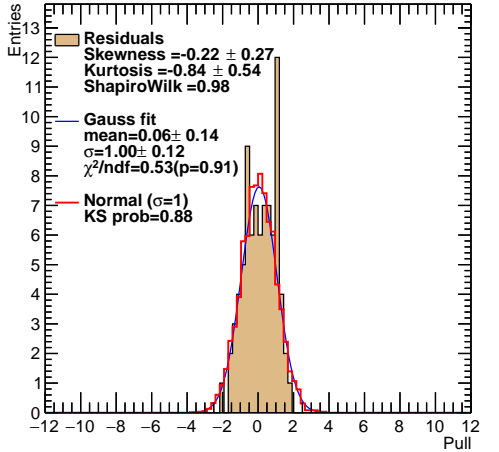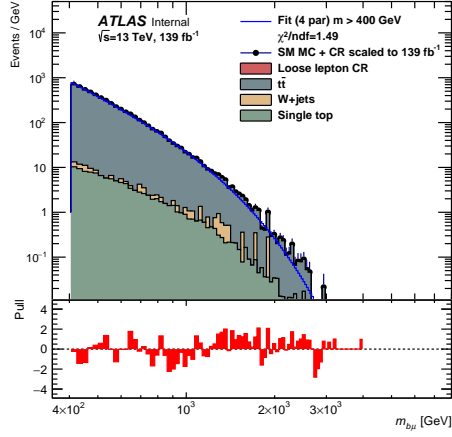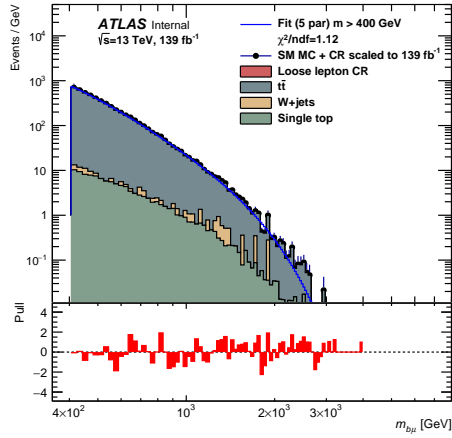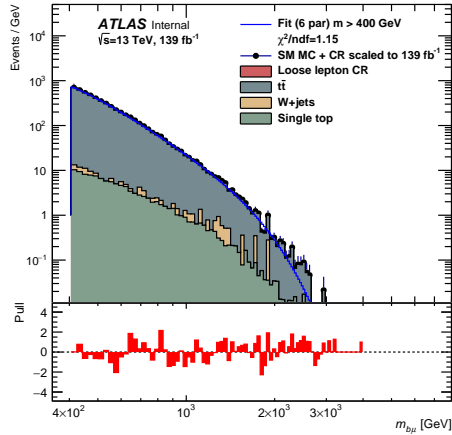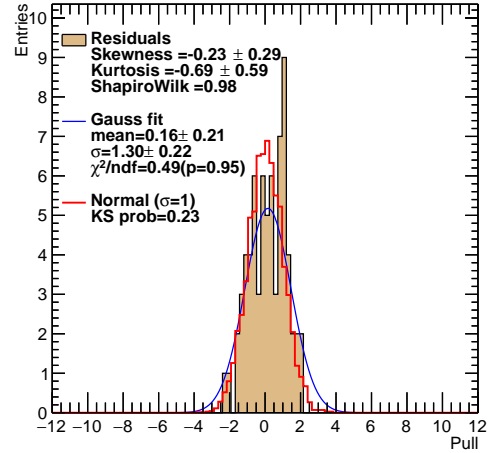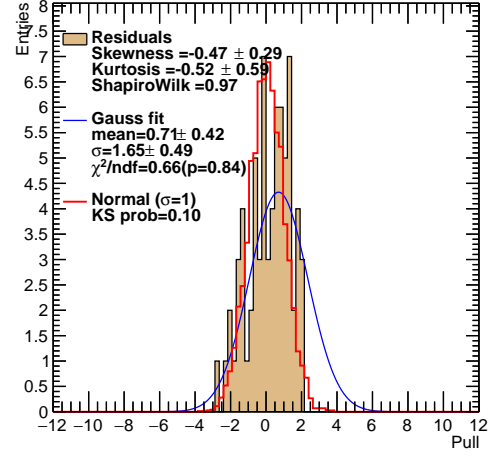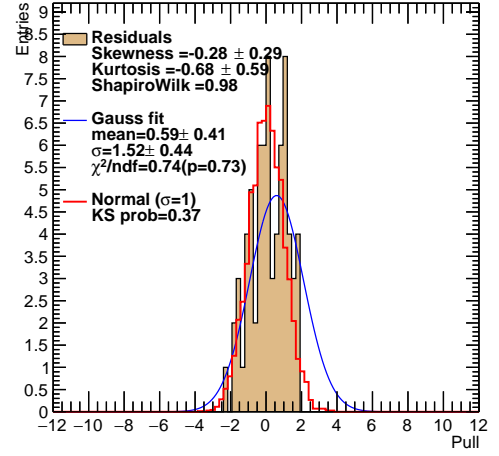

(b) pulls of m$_{jb}$ in p4

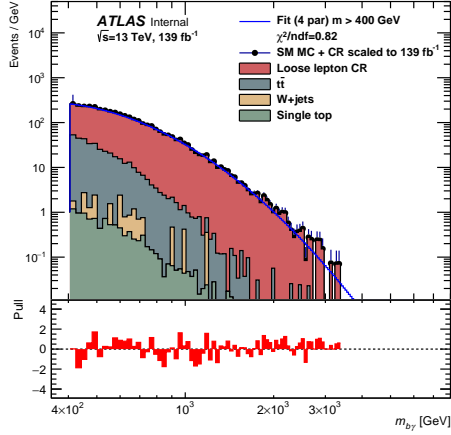(c) m$_{jb}$ using MC+LE-CR, p5

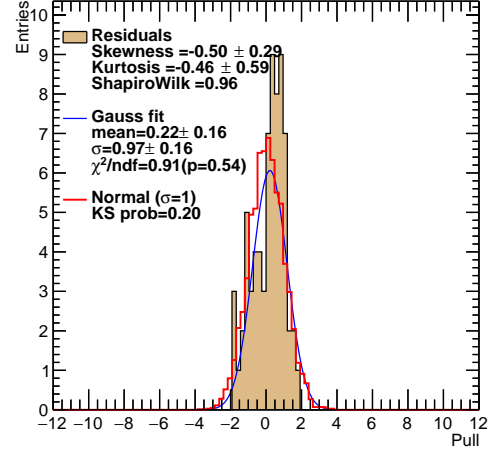(d) pulls of m$_{jb}$ in p5

(e) m$_{jb}$ using MC+LE-CR, p6

(f) pulls of m$_{jb}$ in p6

Figure 60: The m$_{jb}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.

(a) m$_{bb}$ using MC+LE-CR, p4

(b) pulls of m$_{bb}$ in p4

(c) m$_{bb}$ using MC+LE-CR, p5

(d) pulls of m$_{bb}$ in p5

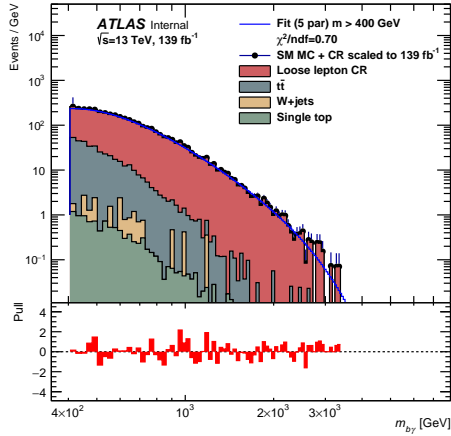(e) m$_{bb}$ using MC+LE-CR, p6

(f) pulls of m$_{bb}$ in p6

Figure 61: The m$_{bb}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.
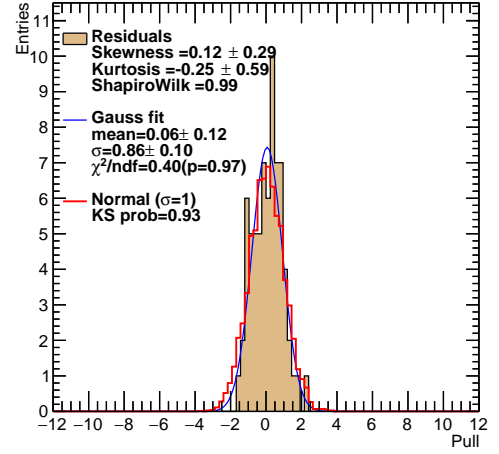
(a) m$_{je}$ using MC+LE-CR, p4

(b) pulls of m$_{je}$ in p4

(c) m$_{je}$ using MC+LE-CR, p5

(d) pulls of m$_{je}$ in p5

(e) m$_{je}$ using MC+LE-CR, p6

(f) pulls of m$_{je}$ in p6

Figure 62: The m$_{je}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.

(a) m$_{j\mu}$ using MC+LE-CR, p4

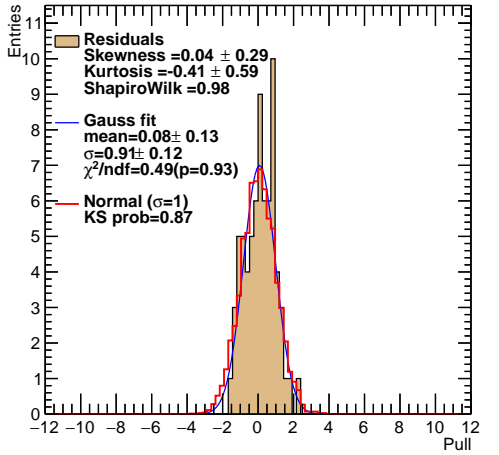(b) pulls of m$_{j\mu}$ in p4

(c) m$_{j\mu}$ using MC+LE-CR, p5

(d) pulls of m$_{j\mu}$ in p5

(e) m$_{j\mu}$ using MC+LE-CR, p6

(f) pulls of m$_{j\mu}$ in p6

Figure 63: The m$_{j\mu}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.

(a) m$_{j\gamma}$ using MC+LE-CR, p4

(b) pulls of m$_{j\gamma}$ in p4

(c) m$_{j\gamma}$ using MC+LE-CR, p5

(d) pulls of m$_{j\gamma}$ in p5

(e) m$_{j\gamma}$ using MC+LE-CR, p6

(f) pulls of m$_{j\gamma}$ in p6

Figure 64: The m$_{j\gamma}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.

(a) m$_{be}$ using MC+LE-CR, p4

(b) pulls of m$_{be}$ in p4

(c) m$_{be}$ using MC+LE-CR, p5

(d) pulls of m$_{be}$ in p5

(e) m$_{be}$ using MC+LE-CR, p6

(f) pulls of m$_{be}$ in p6

Figure 65: The m$_{be}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.

(a) m$_{b\mu}$ using MC+LE-CR, p4

(b) pulls of m$_{b\mu}$ in p4

(c) m$_{b\mu}$ using MC+LE-CR, p5

(d) pulls of m$_{b\mu}$ in p5

(e) m$_{b\mu}$ using MC+LE-CR, p6

(f) pulls of m$_{b\mu}$ in p6

Figure 66: The m$_{b\mu}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.

(a) m$_{b\gamma}$ using MC+LE-CR, p4

(b) pulls of m$_{b\gamma}$ in p4

(c) m$_{b\gamma}$ using MC+LE-CR, p5
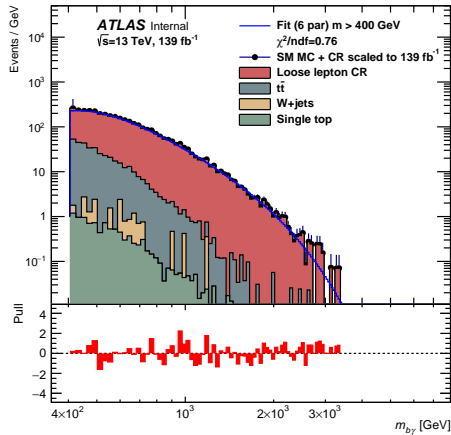
(d) pulls of m$_{b\gamma}$ in p5

(e) m$_{b\gamma}$ using MC+LE-CR, p6

(f) pulls of m$_{b\gamma}$ in p6

Figure 67: The m$_{b\gamma}$ invariant masses with the p4, p5 and p6 fit functions in the BSM region after the 10 pb AR cut is applied. Pulls shown on the right.

The p4 fit shows several failures in the statistical tests. However, the p5 fit function shows good agreement for most of the invariant masses, though there are a few that should be pointed out.

- m_{bb}- Failed 4 tests (mean, sigma, skewness, kurtosis) while the other three tests are good. This is due to problems in MC background that use the LE-CR. Though, this region shows good agreement with the p5 function when using stage-1 10% unblinded data as seen in Appendix APPENDIX J: (only fails two tests).

- $m_{j\mu}$- Failed 4 tests (mean, KS, skewness, kurtosis), while the other three are good. This is mainly because the LE-CR does not contribute to the μ-channel, essentially the fit is only applied to the MC simulations. The stage-1 10% unblinded data in Appendix APPENDIX J: passes all test but one.

6.7 Results

Now that the ARs are defined and a reasonable, well tested fit function has been selected to describe the background hypothesis, results can be obtained by unblinding 100% of the Run 2 data to see if any anomalous resonant signatures are found using the BH strategy. Figure 68 shows two plots using the full 100% unblinded Run 2 data along with the five BSM models (Figure56 shows the analogous 10% of data with scaling).

Figure 68: Distributions of anomaly scores in data and the five benchmark BSM models. (a) shows these BSM models each at the 2 TeV mass hypotheses scaled to the expected events for 140 fb^{-1}. (b) shows these five BSM models at the 6 TeV mass hypothesis also scaled to the expected events for 140 fb^{-1}. The vertical red lines on both show the three defined ARs.

Figure 69 shows the results of the likelihood fit on 100% unblinded Run 2 data for each of the nine invariant masses of interest. This figure is **before** any of the three AR cuts are applied. The largest deviation found using the BH strategy is in the $m_{j\mu}$ channel in the mass range of 0.44-0.48 TeV. The global p-value is 0.057 which corresponds to $Z = 1.5\sigma$. The local significance of this mass range is found to be p-value = 0.00022, corresponding to $Z = 3.5\sigma$. The yellow bands show the fit uncertainty ($\pm\sigma$). Figure 70 shows the BumpHunter results for the fully unblinded Run 2 data within the 10 pb glsar.

Figure 69: BumpHunter results for full unblinded Run 2 data for all nine invariant masses of interest before applying any of the three AR cuts. It also shows the result of the 5p fit function to describe the background hypothesis. The lower panel shows the bin-by-bin fit significances with the largest deviation reported by BumpHunter noted by the vertical dashed lines with its global p-value shown.
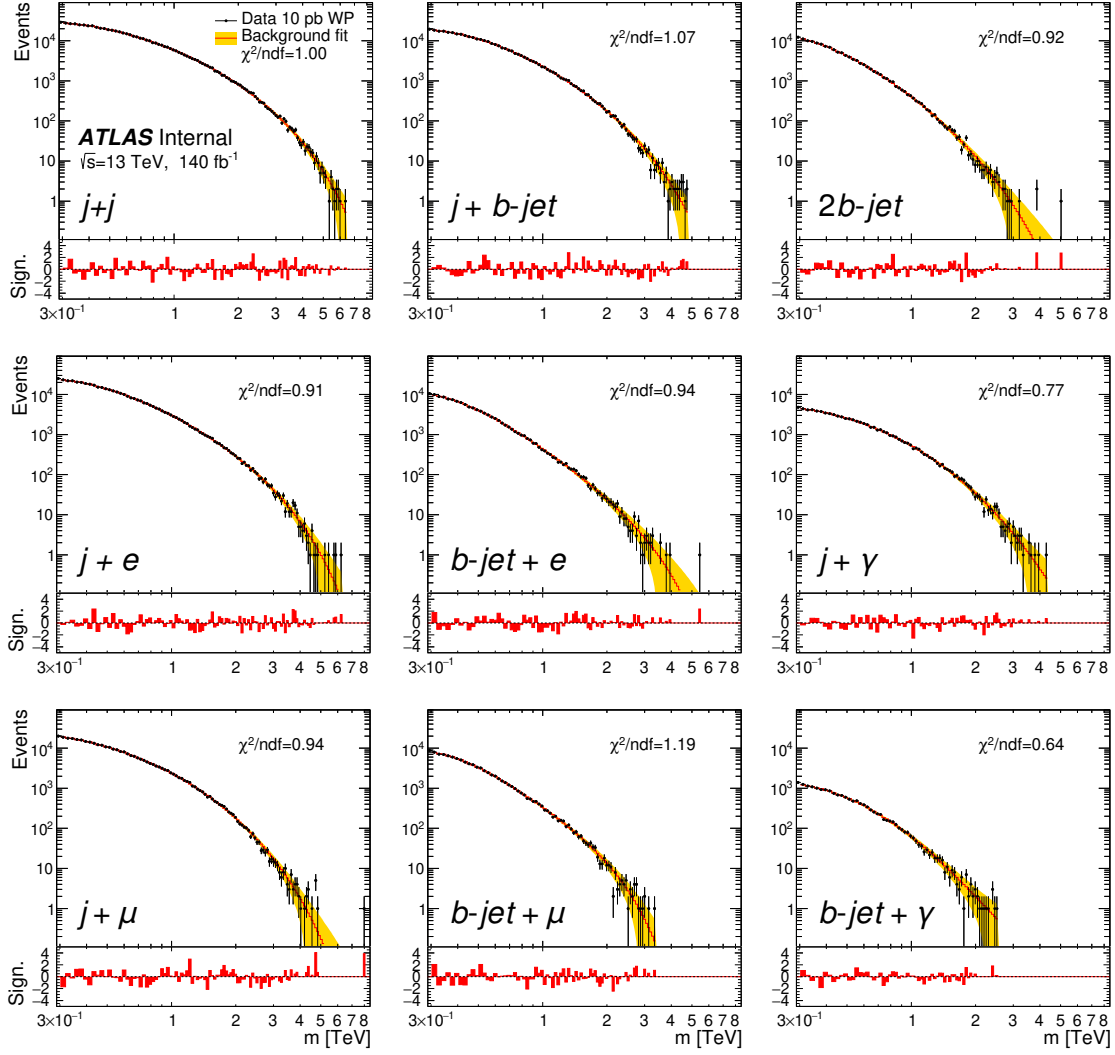
Figure 70: BumpHunter results for full unblinded Run 2 data for all nine invariant masses of interest after applying the 10 pb AR cut. It also shows the result of the 5p fit function to describe the background hypothesis. The lower panel shows the bin-by-bin fit significances with the largest deviation reported by BumpHunter noted by the vertical dashed lines with its global p-value shown.

The largest deviation that's worth discussing can be found in Figure 70 in the $m_{j\mu}$ distribution on the bottom right. This deviance is found near the mass point 4.6-4.8 TeV. This has a local *p*-value of 3.8×10^{-5} corresponding to a significance of $Z = 3.9\sigma$. Accounting for the look-elsewhere effect for the $m_{j\mu}$ distribution, the global *p*-value is 0.013, corresponding to a global significance of $Z = 2.2\sigma$

The metric of discovery sensitivity [34] is briefly discussed in the Appendix APPENDIX G: and is defined in Eq. .1. This value was calculated for each BSM model and all their mass hypothesis. These values are and plotted in Figure 71 in the bin that's associated to its mass. This figure shows the increase of discovery sensitivity after the 10 pb AR cut is applied. Some increases are as large as 200% for some models. Some models, such as SSM, show little (if any) improvement. This is due (in case of SSM) to the fact that the signatures of these models are nearly identical to the SM background. The two plots that involve a photon ($m_{j\gamma}$ and $m_{b\gamma}$) show almost no increase nor decrease. This is because none of the chosen BSM benchmark models require a photon in the final state. There is also a noticeable trend that the discovery sensitivity generally increases as the mass hypotheses of these BSM models increase. This makes sense since high mass particles tend to decay into more anomalous events and therefore will pass the AR cuts.

Figure 71: Bin-by-bin improvement in ΔZ discovery sensitivity after applying the 10 pb AR cut for all nine invariant masses of interest. Discovery sensitivity increases shown for all five benchmark BSM models and all their mass hypothesis.

6.7.1 Limit Setting

Most analyses are non-generic and have a specific BSM model with a range of possible heavy masses that are studied. Limits are then able to be set on these masses generated from the model of interest once the background hypotheses is used to find deviations. This analysis is unique in the case that it is a generic search and only used BSM models for benchmark metrics. It was decided in the end to set limits using Gaussian shapes placed at a large range of possible mass hypotheses.

The limits found in this analysis uses the frequentist method that is based on a profile likelihood

ratio between the fit function, the estimated background, and the possible signature. In the case of upper limits, the statistical model $f(data|\mu, \boldsymbol{\alpha})$, where μ is the parameter representing a signal yield and $\boldsymbol{\alpha}$ are the nuisance parameters. The profile likelihood from the fit function is used to generate different statistics to test alternative signal+background and background hypotheses. To reduce sensitivity, the confidence level (CL) procedure adds an additional test statistic $\bar{q}_{\mu}$ used for upper limits [62]. The *p*-values for the μ(s+b) and μ=0 (background only) hypotheses are analyzed by a large ensemble of pseudo-experiments.

The mean signal Gaussian shapes implemented for limit setting are defined as a hypothetical mass and have values in GeV at:

- 300, 320, 340, 360, 380,

- 500, 530, 560, 590,

- 400, 420, 440, 460, 480,

- 620, 650, 680,

- 710, 750, 790,

- 830, 870,

- 910, 960,

- 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900,

- 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900,

- 3000, 3200, 3400, 3600, 3800,

- 4000, 4200, 4400, 4600, 4800,

- 5000, 5300, 5600, 5900,

- 6200, 6500, 6800,

- 7100, 7500, 7900

The CL level was set to 95% for the observe upper limits on the $\sigma \times BR \times Acc \times Eff$ for Gaussian shaped signal with various widths are seen in Figure 72.
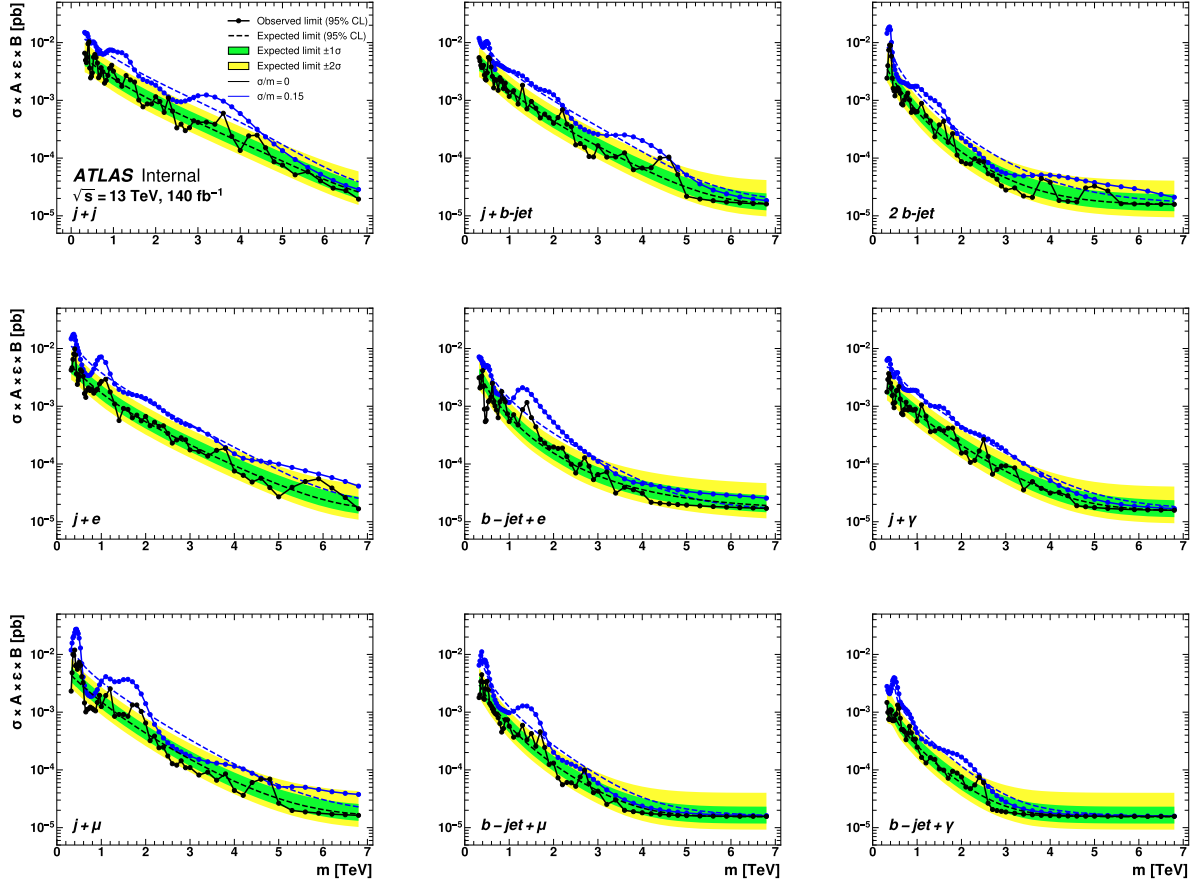
Figure 72: The 95% CL observed upper limits on cross-section times acceptance (A), efficiency (ϵ) and branching ratio (B) for Gaussian signal shapes with various widths. Limits are calculated on events within the 10 pb AR. The $\pm 1\sigma$ and $\pm 2\sigma$ bands are shown for $\sigma_X/m_X = 0$

CHAPTER VII

PRELIMINARY NON-AGNOSTIC BSM MODEL ANOMALY SEARCH

After the success of the agnostic anomaly detection search discussed in Chapter VI, it was decided to try to adapt this technique to look for a specific BSM signal. The frameworks for the anomaly detection were already set in place, all they needed was to be adjusted for a different analysis. This short chapter discusses **very preliminary** plots for another anomaly detection study. Unfortunately, this analysis started on the last part of my degree and I didn't have the time to properly obtain results. Though, the frameworks have been adapted and preliminary anomaly detection studies were conducted. This analysis is expected to be continued by a graduate colleague after my graduation.

7.1 SH → $b\bar{b}b\bar{b}$ Anomaly Detection

This is a search for a massive scalar boson that decays to a light scalar and a Higgs boson in the four b-quark boosted topology. This model expects a massive scalar boson X decay into a lighter scalar boson Y and a SM Higgs boson H through the process X→SH→ $b\bar{b}b\bar{b}$. This search uses proton-proton collision data taken by the ATLAS detector during Run 2 at a CME of 13 TeV corresponding to a luminosity of 138 fb^{-1}. The search is dedicated to a range of mass hypotheses. For the massive X scalar, a range of 0.3-6 TeV is set, and for the lighter Y scalar, the range of 70-5000 GeV. Both the lighter scalar and the SM Higgs boson are Lorentz-boosted and therefore their b quark-antiquark pair are collimated and are reconstructed using a single large-R jet structure. The invariant mass of one pair of the b quark-antiquark pairs is required to be compatible with the Higgs boson of 125 GeV. The approach to find possible resonances and to set limits on these masses will be very similar to the previous anomaly detection technique that was applied in Chapter VI for agnostic

BSM signals.

7.1.1 Event Selection

Since the final topology requires 2 b quark-antiquark pairs, the ideal boosted large-R events contain two double b-tagged large-R jets. The choice of minimum P_T must be sufficient to contain the possibility of having a heaviest massive scalar boson mass hypotheses of 4 TeV. Due to the uniqueness of the required topology, it is crucial to ensure there is enough training statistics for a proper ML model while also not biasing it on signal events. The minimum chosen P_T for the leading large-R jet is taken to be 450 Gev while the second leading large-R jet is required to have a minimum of 250 GeV (P_T(j1(j2))>450(250) GeV), an $|\eta| < 2.0$, and a jet mass of m_j >50 GeV. These large-R jets undergo a b-tagging criteria using the FTAG Xbb2020v3 tagger at the 70% WP. This tagger is much like the DL1d tagger than can identify between b-, c-, and light-quarks up to a certain efficiency, but instead of a single quark, the Xbb tagger can identify large-R jets that may contain two b-jets up to a certain efficiency. These events that contain a single large-R jet that is double b-tagged as a 1bb-tagged and events with two large-R jets that are both double-tagged as 2bb-tagged (events may contain other non-b-tagged jets). Figure X shows the cutflow and the chosen NN training sample statistics. Using events that just contain a single 1bb-tagged large-R jet would lead to insufficient statistics to train a proper AE model, therefore the training sample was chosen using the P_T(j1(j2))>450(250) GeV cut.

7.2   Event Representation

Following the example in the anomaly detection analysis, a modified RMM representation was chosen as the input to the deep-learning autoencoder (AE) architecture. Since the events only contain x amount of large-R jets and y amount of double b-tagged large-R jets, the amount of input variables allowed within this representation decreases from the original layout. It was chosen to allow up to 5 large-R jets and up to 4 double b-tagged large-R jets. This layout contains a total of 100 input variables for the AE (including one row and column of zeroes separating the large-R
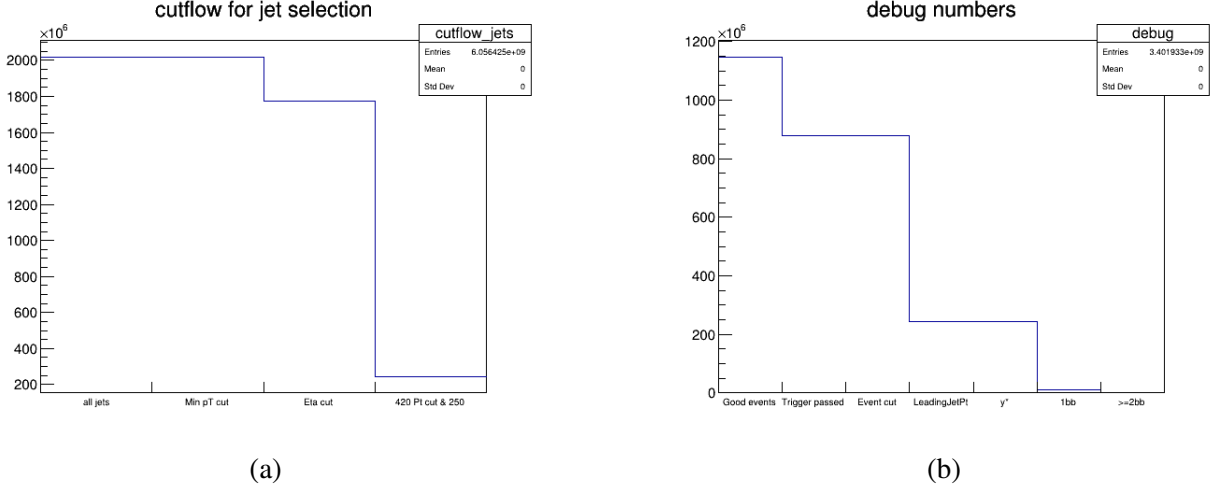
Figure 73: (a) Shows the cutflow for Run 2 events. The last cut is on (j1(j2))>450(250) GeV. This cut equates to about 240M events. (b) Shows this 240M events that make the base cut and then shows how many are 1bb-tagged and 2bb-tagged. This equates to 1M events for 1bb-tagged and 6.2K for 2bb-tagged.

jets and the double b-tagged large-R jets). Figure X shows a diagram of this representation. The information layout is the same as the previous version discussed in Section 6.4. Figure X shows two examples of this RMM representation. The left figure shows a single event containing three large-R jets and no double b-tagged large-R jets. The right plot shows all the events stacked into a single matrix. This plot reveals there are events with 5 large-R jets.
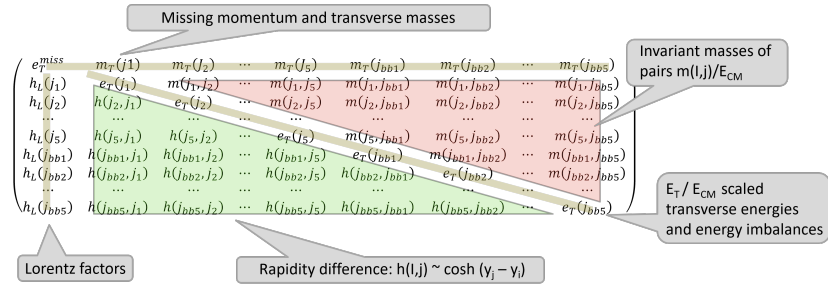
Figure 74: The Rapidity Mass Matrix representation for the SH → 4b analysis. This layout allows up to 5 large-R jets and 4 double b-tagged large-R jets. This topology equates to 100 variables used for training the Autoencoder (included a row and column of zeroes).

7.2.1 Autoencoder Training

The preliminary AE architecture can be seen in Figure 20. Just as previously discussed, this architecture contains an encoder, latent layer, and a decoder. This architecture compresses data from

150
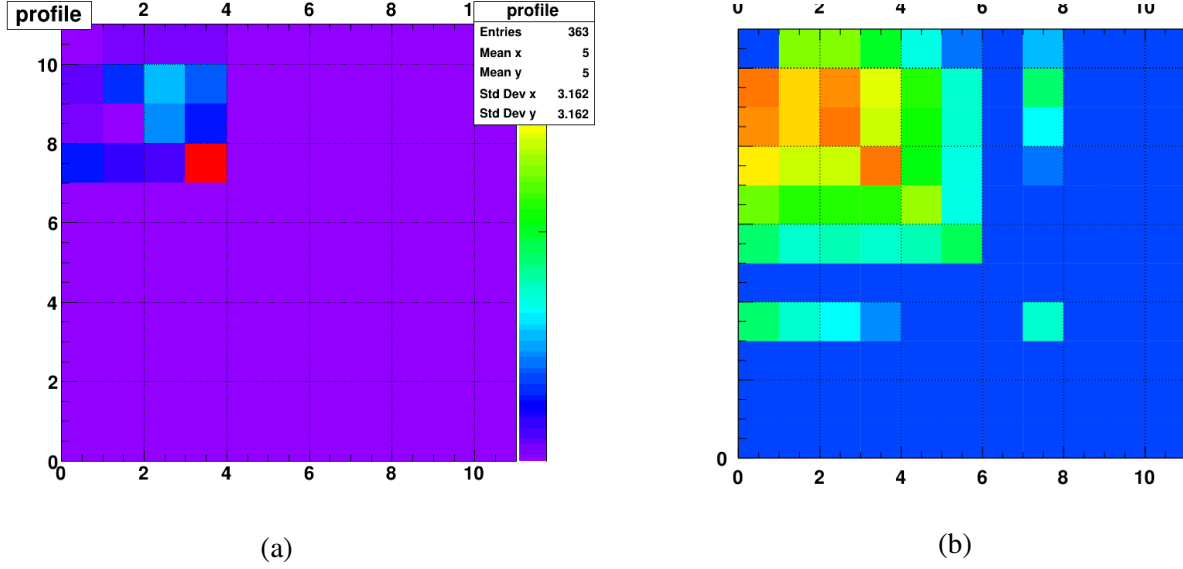
Figure 75: (a) Shows a single event that contains three large-R jets using this new RMM representation. (b) Shows all the events stacked into a single RMM. This plot shows there are events that contain 5 large-R jets. The row and column for the second double b-tagged large-R jet show to be almost empty, this is due to the fact that there are only 6.2K with respect to a total of 240M events shown in the first few rows and columns.

the input into the latent, then decompresses it using the decoder. The loss value calculated by the loss function is recorded and used for the anomaly score as the form of log(*loss*). This model was training using 1% of Run 2 large-R jet data, equating to 2.4M events with a 7:3 ratio for training and validation respectively. There requires many more studies in order to optimize this architecture but this was used for preliminary loss distribution studies.

7.3   Preliminary Loss Distribution Studies

The data sample contains a total of 240M events consisting of large-R jets with that made the cut of P_T(j1(j2))>450(250) GeV. This dataset consists of the entire Run 2 data-taking campaign that was taken over the years of 2015-2018. This dataset was then split into their respective years and plotted in order to see if there are any year dependence due to pile-up conditions. This can be seen in Figure 77 (the ratio plot needs to be adjusted).

The distribution in Figure 78 shows cascading of four peaks. This is due to the amount of objects allowed in the input representation. Figure X points out the origination of these peaks. The far most
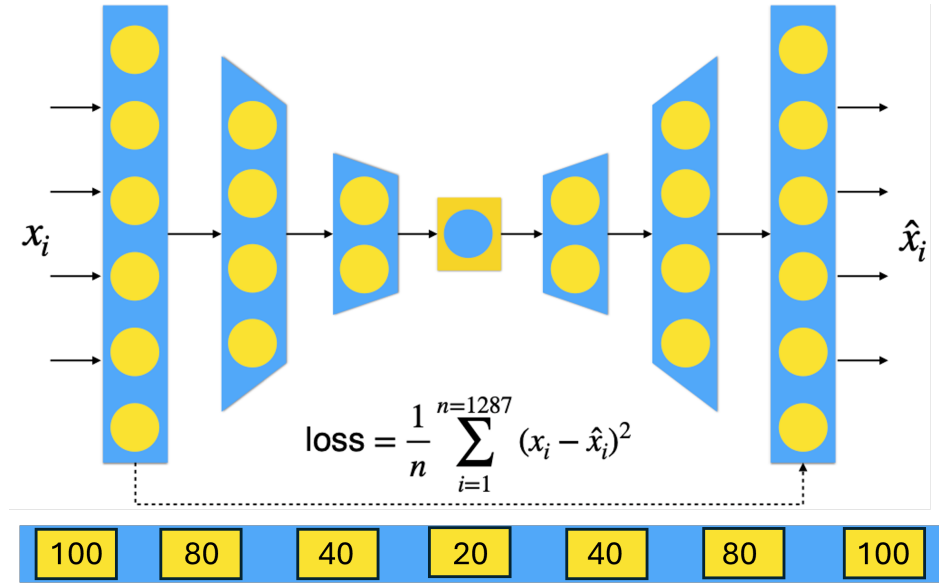
Figure 76: Autoencoder architecture diagram for preliminary studies for the SH→4b analysis. The first part of this architecture consists of an encoder that compresses data into the latent layer, the decoder then decompresses the data and attempts to reconstruct the original input. The number of neural nodes are shown on the bottom.

left peak is the least anomalous events which contain two large-R jets that are not double b-tagged. The second most left peak are events that contain three large-R jets. The third peak is then events that contain at least one large-R jet and one 1bb-tagged large-R jet. The final peak on the right is the events that contain 2bb-tagged large-R jets. There are other jets with odd combinations that do exist as can see in the validation RMM plot on the right in Figure 75 which are contained in or between these peaks.
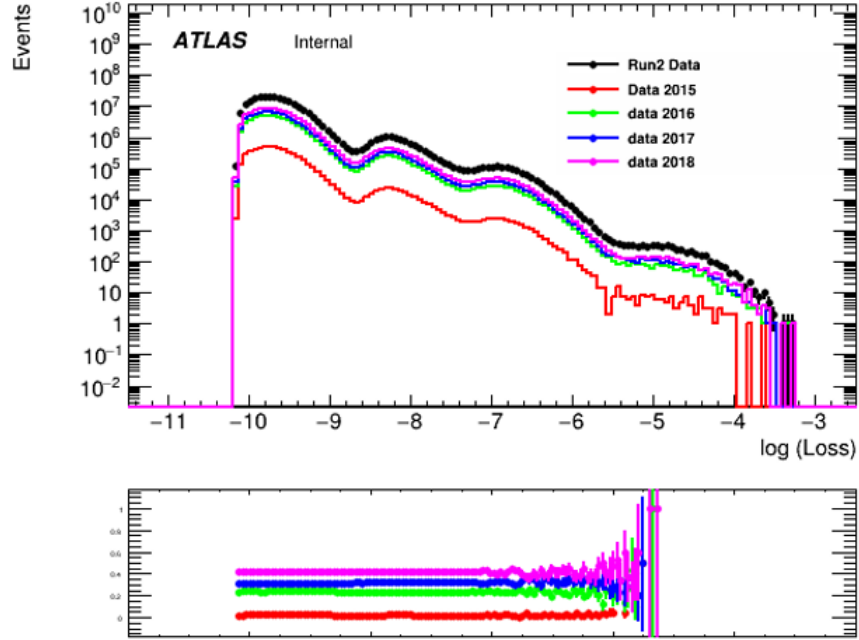
Figure 77: The log(*loss*) distribution for all of Run 2 and its data-taking years. No yearly dependence is observed. The ratio plot needs to be adjusted

There are six mass hypotheses for the large scalar boson and the light scalar boson that was chosen to help with benchmarking the AE model. These six mass hypotheses are:

1. Large Scalar Mass X: **1 TeV**, Light Scalar Mass S: **500 GeV**

2. Large Scalar Mass X: **3 TeV**, Light Scalar Mass S: **750 GeV**

3. Large Scalar Mass X: **300 GeV**, Light Scalar Mass S: **70 GeV**

4. Large Scalar Mass X: **6 TeV**, Light Scalar Mass S: **1 TeV**

5. Large Scalar Mass X: **6 TeV**, Light Scalar Mass S: **5 TeV**

6. Large Scalar Mass X: **750 GeV**, Light Scalar Mass S: **250 GeV**

The loss distributions of these six models along with data are seen in Figure 79. These models are not scaled to their expected event count from Run 2 luminosity, but they do demonstrate the separation power of the AE.
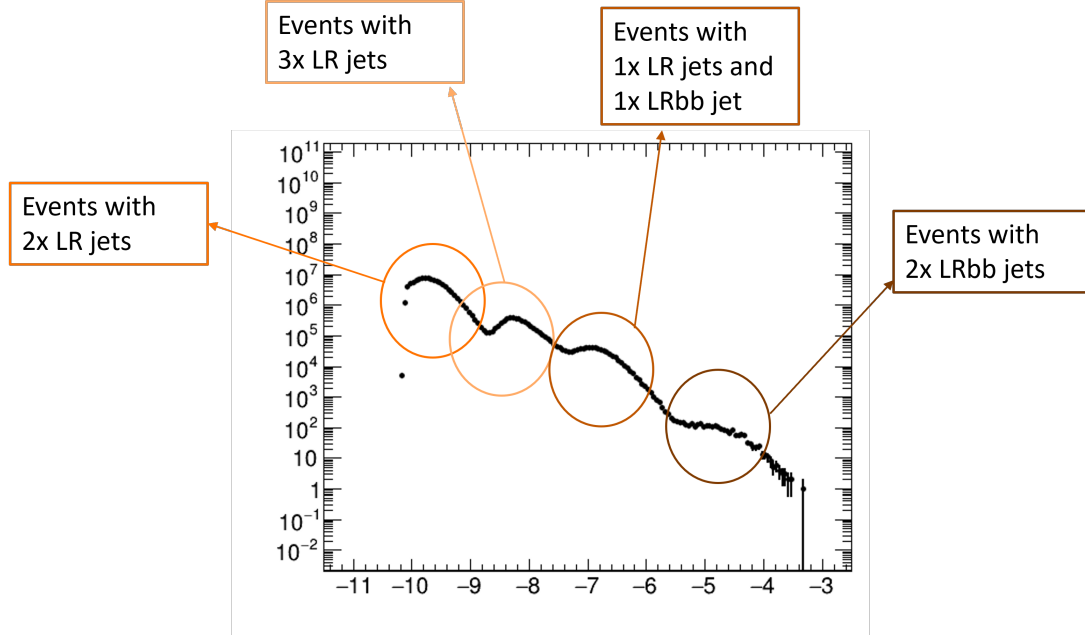
Figure 78: Diagram of data log(*loss*) distribution showing the origination of its four peaks.

7.3.1 Preliminary Anomaly Region Definition

The previous anomaly detection analysis used a calculated BSM event count defined from the benchmark BSM models used within the studies. That is one approach to define the anomaly region. The two definitions that are currently used (subject to change) are based on the double b-tagged event counts. The first AR is of those that contain a single 1bb-tagged large-R jet. A second region is based on the event count of those that contain 2bb-tagged large-R jets. The ARs that contain the correlated event counts can be seen in Figure X as a red line and pink dashed line.

154
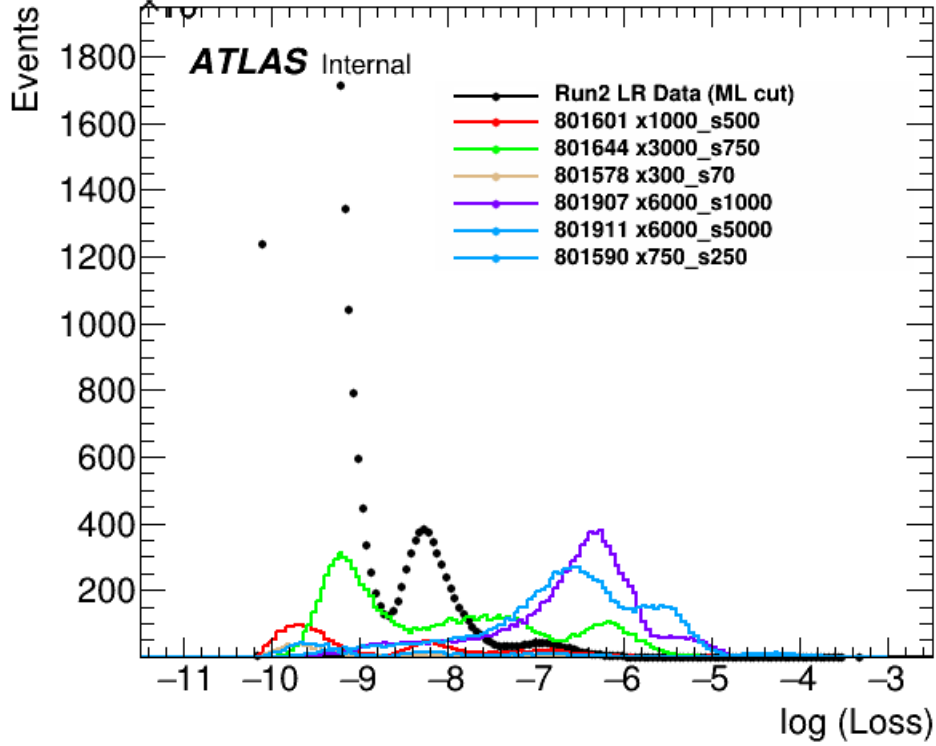
Figure 79: Log(*loss*) for Run 2 data and six mass hypothesis for the large scalar boson X and the lighter scalar boson S. The BSM events are not scaled to their expected event count with respect to the luminosity. The y-axis is not log scaled.

7.4 Conclusion

These studies are still in the preliminary stage and should not be taken as any final result. This work is showing promise and should end in an interesting result, especially when it gets to set limits on the mass hypotheses.

There needs to be many more studies on the finding the optimized AE architecture. Also, is the current chosen RMM representation the best choice? Is it preferred to choose a topology that won't discriminate between the number of objects within the event? The BSM models originally required to have 2bb-tagged large-R jets, so maybe this discrimination may work out best. Should the AR cuts be based on the event counts of 1bb and 2bb-tagged large-R jets or should it be based on the event counts of the BSM models themselves. MC background events were simulated but not yet studied for this analysis which are crucial to start looking into the fit procedure. These are but a few tasks that need to be studied.
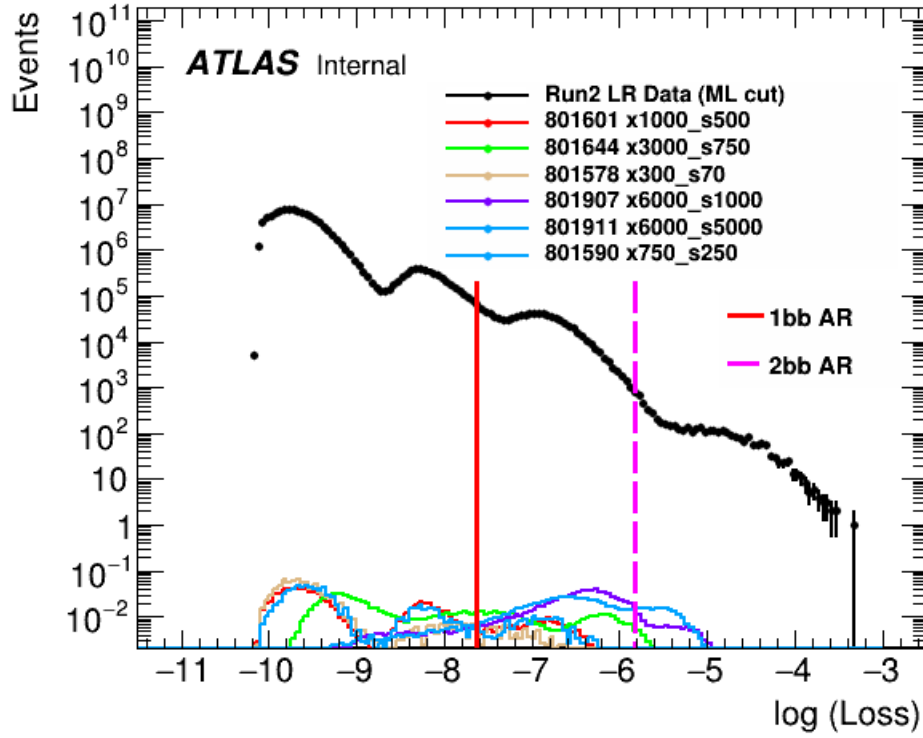
Figure 80: Log(*loss*) for Run 2 data and the six mass hypotheses. Two AR cuts are used to using the event count for 1bb-tagged large-R jets and 2bb-tagged large-R jets. BSM models integrals are scaled to 1. The 1bb-tagged AR is seen as a red line and the 2bb AR is seen as a pink dashed line.

CHAPTER VIII

CONCLUSION

The SM of particle physics has proven to be one of the most successful theories that has ever been developed, standing firm through decades of research while being solidified more through the discovery of the Higgs boson. Even though its success is unprecedented, there are gaps that need to be filled in order to describe phenomena such as gravity and dark matter. As the field of High Energy Physics grows to help explain and fill these gaps, new detectors, techniques and tools are needed to be developed. This dissertation presents two studies that help this ever needed growth, along with a proposed non-generic anomaly detection search that is in its infant stage. One presents preliminary studies of particle identification machine learning tools using simulated data of the ATLAS detector with its upgrade that is being implemented in 2029 for the evolution of the HL-LHC. The second study shows a full analysis that was the first of its kind using unsupervised machine learning for anomaly detection. A non-generic anomaly detection analysis with very preliminary plots is also discussed.

The DL1d tagger developed for the ATLAS Upgrade for the HL-LHC shows promising progress. This is known to be preliminary due to the fact that the training statistics, which is crucial for its effectiveness, has a magnitude less than the current ATLAS standard tagger for Run 2. Regardless of this fact, the trained DL1d tagger for Upgrade shows to be as effective at higher working points while also allowing for the tagging of particles at $|\eta| \geq 2.4$ due to the new implemented ATLAS geometry. The DL1d tagger is in the works of being replaced by a tagger based on a graph neural network architecture and is planned to the standard used tagger once the Upgrade is complete in 2029. Therefore, as of right now, there is no plan to retrain the DL1d tagger for Upgrade and the one that is trained in this dissertation will be used as the industry baseline.

The final analysis shows a novel anomaly detection technique using unsupervised machine learning for anomaly detection training on 1% of Run 2 data. The deep-learning architecture of an AE is used with its reconstruction loss defining the anomaly score. From here, anomaly regions are defined by using the anomaly score as cut on SM events. Bump hunting strategies are discussed and implemented on nine different di-object invariant mass distributions in order to find any deviations within these anomaly regions that may hint at signatures of new physics. The largest deviation found is within the $m_{j\mu}$ distribution in the mass range of 4.6-4.8 TeV. This deviation resulted in a local significance of 3.9σ and a global significance of 2.2σ. Frequentist limits were taken for Gaussian shape signals with means at a ranges defined by benchmark BSM models.

The beginning of a non-generic anomaly detection search using the strategy discussed in this analysis for a specific BSM model of a heavy scalar boson is briefly mentioned. The plots shown are very preliminary. A fellow graduate student is expected to continue this analysis in hopes it may discover hints of new physics.

REFERENCES

[1] J. Faltz A. Rosenfeld, *Sequential operations in digital picture processing*, J. ACM (1966), no. 13, 471–494.

[2] Q. R. Ahmad et al, *Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory*, Phys. Rev. Lett. **89** (2002), 011301.

[3] ATLAS Collaboration Et al., *Electron and photon energy calibration with the atlas detector using 2015-2016 lhc proton-proton collision data*, In: JINST (2019), no. 14.03, P03017.

[4] ———, *Electron reconstruction and identification in the atlas experiment using the 2015 and 2016 lhc proton–proton collision data at $\sqrt{s}$ = 13 tev*, In: Eur. Phys. **J** (2019), no. C 79, 639.

[5] B. M. G. Altarelli and M. Ruiz-Altaba, *Searching for new heavy vector bosons in pp colliders*, Z. Phys. **C 45** (1989), no. 109.

[6] Et al. ATLAS collab., *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s}$ = 8 tev using the atlas detector*, In: Eur. Phys. (2016), no. J. C 76.11, 581.

[7] et al. ATLAS collab., *Forward jet vertex tagging using the particle flow algorithm*, In: Tech. rep. (2019), no. ATL-PHYS-PUB-2019-026.

[8] et al. ATLAS Collaboration, *Optimisation of the atlas b-tagging performance for the 2016 lhc run*, ATL-PHYS-PUB-2016-012 (2016).

[9] Et al. ATLAS Collaboration, *Performance of the atlas track reconstruction algorithms in dense environments in lhc run 2*, In: Eur. Phys. J. C (2017), no. 77.10, 673.

[10] et al. ATLAS Collaboration, *Atlas b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at surds = 13 tev*, In. Eur. Phys (2019), no. J. C 79, 970.

[11] ——, *Deep sets based neural networks for impact parameter flavour tagging in atlas*, ATL-PHYS-PUB-2020-014 (2020).

[12] Et al. ATLAS Collaboration, *Operation of the atlas trigger system in run 2*, JINST (2020), no. 15.10, P10004.

[13] Et al. ATLAS collaboration, *Optimisation of large-radius jet reconstruction for the atlas detector in 13 tev proton-proton collisions*, In: Eur. Phys. J. C (2021), no. 81.4, 334.

[14] et al. ATLAS Collaboration, *Graph neural network jet flavour tagging with the atlas detector*, ATL-PHYS-PUB-2022-027 (2022), 5–7.

[15] ——, *Neural network jet flavour tagging with the upgraded atlas inner tracker detector at the high-luminosity lhc*, ATL-COM-PHYS-2022-934 (2022).

[16] N. Berger, *Recommendations for the modeling of smooth backgrounds*, tech. rep. ATL-COM-PHYS-2020-418 (2020).

[17] G. Ingelman Bo Andersson, G. Gustafson and T. Sjöstrand, *Parton fragmentation and string dynamics*, Phys. Rept. (1983), no. 97, 31–145.

[18] et al. Béjar Alonso, *High-luminosity large hadron collider (hl-lhc): Technical design report*, CERN Archives **10** (2020).

[19] M. Chala and M. Spannowsky, *Behavior of composite resonances breaking lepton flavor universality*, Phys. Rev. **D** (2018), no. 98.

[20] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π decay of the k_2^0 meson*, Phys. Rev. Lett. **13** (1964), 138–140.

[21] ATLAS Collab., *Performance of jet substructure techniques for large-r jets in proton-proton collisions at $\sqrt{s}$ = 7 tev using the atlas detector*, JHEP (2013), no. 09, 076.

[22] ———, *Atlas pythia 8 tunes to 7 tev data*, ATL-PHYS-PUB-2014-021 (2014), no. 191.

[23] ———, *The anti-kt jet clustering algorithm*, In: Eur. Phys. (2017), no. J. C 77.7, 466.

[24] ———, *In situ calibration of large-radius jet energy and mass in 13 tev proton-proton collisions with the atlas detector*, In: Eur. Phys. (2019), no. J. C 79.9, 135.

[25] ATLAS collab., *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s}$ = 13 tev with the atlas detector*, In: Eur. Phys. (2021), no. J. C 81.8, 689.

[26] ATLAS Collab., *Optimisation of large-radius jet reconstruction for the atlas detector in 13 tev proton-proton collisions*, In: Eur. Phys. (2021), no. J. C 81.4, 334.

[27] ATLAS collab. Et al., *Topological cell clustering in the atlas calorimeters and its performance in lhc run 1*, In: Eur. Phys. (2017), no. J. C 77, 639.

[28] ALICE Collaboration, *The alice experiment at the cern lhc*, JINST **3** (2008), S08002.

[29] ATLAS Collaboration, *The atlas experiment at the cern large hadron collider*, JINST **3** (2008), S08003.

[30] ———, *Technical design report for the atlas inner tracker strip detector*, CERN, tech. rep. (2017).

[31] ATLAS collaboration, *Electron and photon performance measurements with the atlas detector using the 2015-2017 lhc proton-proton collision data*, JINST (2019), no. 14.12, P12006.

[32] CMS Collaboration, *The cms experiment at the cern lhc*, JINST **3** (2008), S08004.

[33] LHCb Collaboration, *The lhcb detector at the lhc*, JINST **3** (2008), S08005.

[34] G. Cowan, *Discovery sensitivity for a counting experiment with with background uncertainty*, (2012).

[35] ATLAS Collab. et al., *The simulation principle and performance of the atlas fast calorimeter simulation fastcalosim*, Tech. Rep. ATL-PHYS-PUB-2010-013 (2010).

[36] ATLAS Collab et. al, *A high-granularity timing detector (hgtd) in atlas : Performance at the hl-lhc*, ATL-LARG-PROC-2018-003 (2018), 6.

[37] _____ , *Electron reconstruction and identification in the atlas experiment using the 2015 and 2016 lhc proton–proton collision data at $\sqrt{s}$ = 13 tev*, In: Eur. Phys. (2019), no. J. C 79, 693.

[38] ATLAS Collab. et al., *Search for dijet resonances in events with an isolated charged lepton using $\sqrt{s}$ = 13 tev proton-proton collision data collected by the atlas detector*, JHEP 06 (2020).

[39] ATLAS Collab et. al, *Neural network jet flavour tagging with the upgraded atlas inner tracker detector at the high-luminosity lhc*, ATL-PHYS-PUB-2022-047 (2022), 3–4.

[40] ATLAS Collab. et al., *Search for new phenomena in multi-body invariant masses in events with at least one isolated lepton using 13 tev proton-proton collision data collected by the atlas detector*, tech. rep. ATLAS-CONF-2022-048 (2022).

[41] ATLAS Collaboration et al., *Search for dijet resonances in events with an isolated lepton using $\sqrt{s}$ = 13 tev proton-proton collision data collected by the atlas detector*, tech. rep. ATLAS-CONF-2018-015 (2018).

[42] _____ , *Search for low-mass dijet resonances using trigger-level jets with the atlasdetector in p p collisions at $\sqrt{s}$ = 13 tev*, arXiv (2018), no. 1804.03496.

[43] _____ , *Electron and photon performance measurements with the atlas detector using the 2015-2017 lhc proton-proton collision data*, In: JINST (2019), no. 14.12.

[44] Charles R. Harris et al., *Array programming with numpy*, In: Nature 585.7825 (2020), 357–362.

[45] Daniel Hay Guest et al., *lwtnn/lwtnn*, Version v2.9 (2019).

[46] Enrico Bothmann et al., *Event generation with sherpa 2.2*, In: SciPost Phys. (2019), no. 7.3, 034.

[47] F. Pedregosa et al., *Scikit-learn: Machine learning in python*, (2011), no. 12, 2825–2830.

[48] Felipe Pezoa et al., *Foundations of json schema*, In: Proceedings of the 25th International Conference on World Wide Web **International World Wide Web Conferences Steering Committee** (2016), 263–273.

[49] François Chollet et al., *keras*, (2015).

[50] Johannes Bellm et al., *Herwig 7.0/herwig++ 3.0 release note*, In: Eur. Phys. (2016), no. J. C 76.4, 196.

[51] K. S. Agashe et al, *Lhc signals from cascade decays of warped vector resonances*, JHEP **05** (2017), no. 078.

[52] M. Bähr et al, *Herwig++ physics and manual*, In: Eur. Phys. (2008), no. J. C 58, 639.

[53] M. Tanabashi et al., *Forward jet vertex tagging using the particle flow algorithm*, In: Phys. Rev. **D** (2018), no. 98, 03001.

[54] M. Zaheer et al, *Deep sets*, CoRR (2017), 2,5.

[55] Martín Abadi et al., *Tensorflow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org (2015).

[56] R. Alemany-Fernandez et al., *Operation and configuration of the lhc in run 1*, Tech. rep. (2013).

[57] S. Agostinelli et al., *Geant4 λ a simulation toolkit*, Nucl. Instrum. Meth. **A** (2003), no. 506, 250.

[58] T. Cornelissen et al, *Concepts, design and implementation of the atlas new tracking (newt)*, Tech. rep. **ATL-SOFT-PUB-2007-007** (2007).

[59] L. Evans and P. Bryant, *Lhc machine*, JINST **3, S08001** (2008).

[60] R. Frühwirth, *Application of kalman filtering to track and vertex fitting*, Nucl. Instrum. Methods **A** (1987), no. 262, 444.

[61] Et al. G. Aad, *Muon reconstruction and identification efficiency in atlas using the full run 2 pp collision data set at $\sqrt{s}$ = 13 tev*, In: EUR. Phys. **J. C 81.7** (2021), 578.

[62] E. Gross G. Cowan, K. Cranmer and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physics Journal **ISSN: 1434-6052** (2011), no. C 71.

[63] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961), 579–588.

[64] Mark I. Grossman, *John dalton and the london atomists*, Notes and Records of the Royal Society of London **68(4)** (2014), no. 68, 339–356.

[65] The HDF Group, *Hierarchical data format version 5*, (2000-2010).

[66] Laura Reina Heribertus B. Hartanto, Barbara Jäger and Doreen Wackeroth, *Higgs boson production in association with top quarks in the powheg box*, In: SciPost Phys. **D** (2019), no. 91.9, 094003.

[67] W. Herr and B. Muratori, *Concept of luminosity*, In: CERN Accelerator School at DESY Zeuthen: Accelerator Physics (2003), 364.

[68] Peter W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964), 508–509.

[69] J. D. Hunter, *Matplotlib: A **2**d graphics environment*, In: Computing in Science and Engineering (2007), no. 9.3, 90–95.

[70] Stefan Höche, *Introduction to parton-shower event generators*, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier **Amplitudes for Colliders** (2015), 235–295.

[71] J. Illingworth and J. Kittler, *A survey of the hough transform. comput. vis. gr. image process*, issn: 0734-189X (1988), no. 44, 87.

[72] K. Ying J. Luo and J. Bai, *Savitzky-golay smoothing and differentiation filter for even number data*, Signal Processing **ISSN: 0165-1684** (2005), no. 85.

[73] Frank Krauss Jan-Christopher Winter and Gerhard Soff, *A modified cluster-hadronization model*, Eur. Phys J. C (2004), no. 36, 381–395.

[74] et al. Johan Alwall, *Madgraph/madevent v4: The new web generation*, JHEP (2007), no. 09, 028.

[75] Thomas Junk and Louis Lyons, *Reproducibility and Replication of Experimental Particle Physics Results*, Harvard Data Science Review **2** (2020), no. 4, https://hdsr.mitpress.mit.edu/pub/1lhu0zvn.

[76] S. Hong K. Agashe, P. Du and R. Sundrum, *Flavor universal resonances and warped gravity*, JHEP **01** (2017), no. 016.

[77] G. P. Salam M. Cacciari and G. Soyez, *The anti-kt jet clustering algorithm*, In: JHEP (2008), no. 04, 063.

[78] T. Maskawa M. Kobayashi, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49(2)** (1973), 652–657.

[79] M. Magg and C. Wetterich, *Neutrino mass problem and gauge hierarchy*, Phys. Rev. **B** (1981), no. 98.

[80] Dirk Merkel, *Docker: lightweight linux containers for consistent development and deployment*, In: Linux journal (2014), no. 2014.239, 2.

[81] E. Noether, *Invariante variationsprobleme*, Math-phys. Klasse (1918), no. Seite 235-157.

[82] The pandas development team, *pandas-dev/pandas: Pandas*, Version latest (2020).

[83] J. Pequenao and P. Schaffner, *How atlas detects particles: diagram of particle paths in the detector*, (2013).

[84] A. Pich, *The standard model of electroweak interactions*, arXiv:hep-ph/0502010 (2004), no. IFIC/05-13, FTUV/05-0201.

[85] Guido Van Rossum and Fred L. Drake, *Python 3 reference manual*, Scotts Valley, CA: CreateSpace (2009).

[86] Andrei D. Sakharov, *Violation of cp invariance, c asymmetry, and baryon asymmetry of the universe*, Soviet Physics Uspekhi **34** (1991-05), 392.

[87] Abdus Salam and JC Warc, *Electromagnetic and weak interactions*, Physics Letters **13** (1964), 168–174.

[88] et al. Torbjörn Sjöstrand, *An introduction to pythia 8.2*, In: Comput. Phys. Commun (2015), no. 191, 159.

[89] Steven Weinberg, *A model of leptons*, Physical Review Letters **19** (1967), no. 21.

[90] J. Wenninger, *Operation and configuration of the lhc in run 2*, Tech. rep. (2019).

[91] R.L. Workman, *Review of particle physics*, PTEP **2022** (2022), cit on pp. 23,25.

[92] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental test of parity conservation in beta decay*, Phys. Rev. **105** (1957), 1413–1415.

[93] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954), 191–195.

[94] F. Yu, *A z' model for the cdf dijet anomaly*, Phys. Rev. **D83** (2011), no. 094028.

APPENDICES

APPENDIX A: GN1 for Upgrade

The GN1 high-level tagger is the newest and the first of the graph neural network architectures lineage. This tagger was designed and studied for Run 2, Run 3, and Run 4 (Upgrade) of the ATLAS detector. This tagger's architecture was briefly described in Section 4.6.4. The input variables for this tagger can be seen in Table 24. An even newer version of GN1 tagger has been developed and currently being studied for its correlated scale factors that utilizes auxiliary tasks. This newer version is called the GN2 tagger.

The GN1 tagger was studied using the same Upgrade samples listed in Table 13 and had its performance comapred to the DL1d tagger that was trained for this thesis that can be found in Chapter V. The following plots show the performance of the GN1 tagger to the DL1d and MV2 high-level taggers. The top plots show c-jet rejection and ligh-jet rejection using the $t\bar{t}$ sample from Table 13. The middle plots show all three taggers performances using the Z' samples. The results show the clear superiority using the graph neural networks, increasing b-tagging performance within c-jet background and light-jet background of up to 2.5x. The last two plots show the performances on jetes with the new eta range that is available from the ITk detector. Again, the GN1 tagger massively outperforms the DL1d and MV2 taggers.
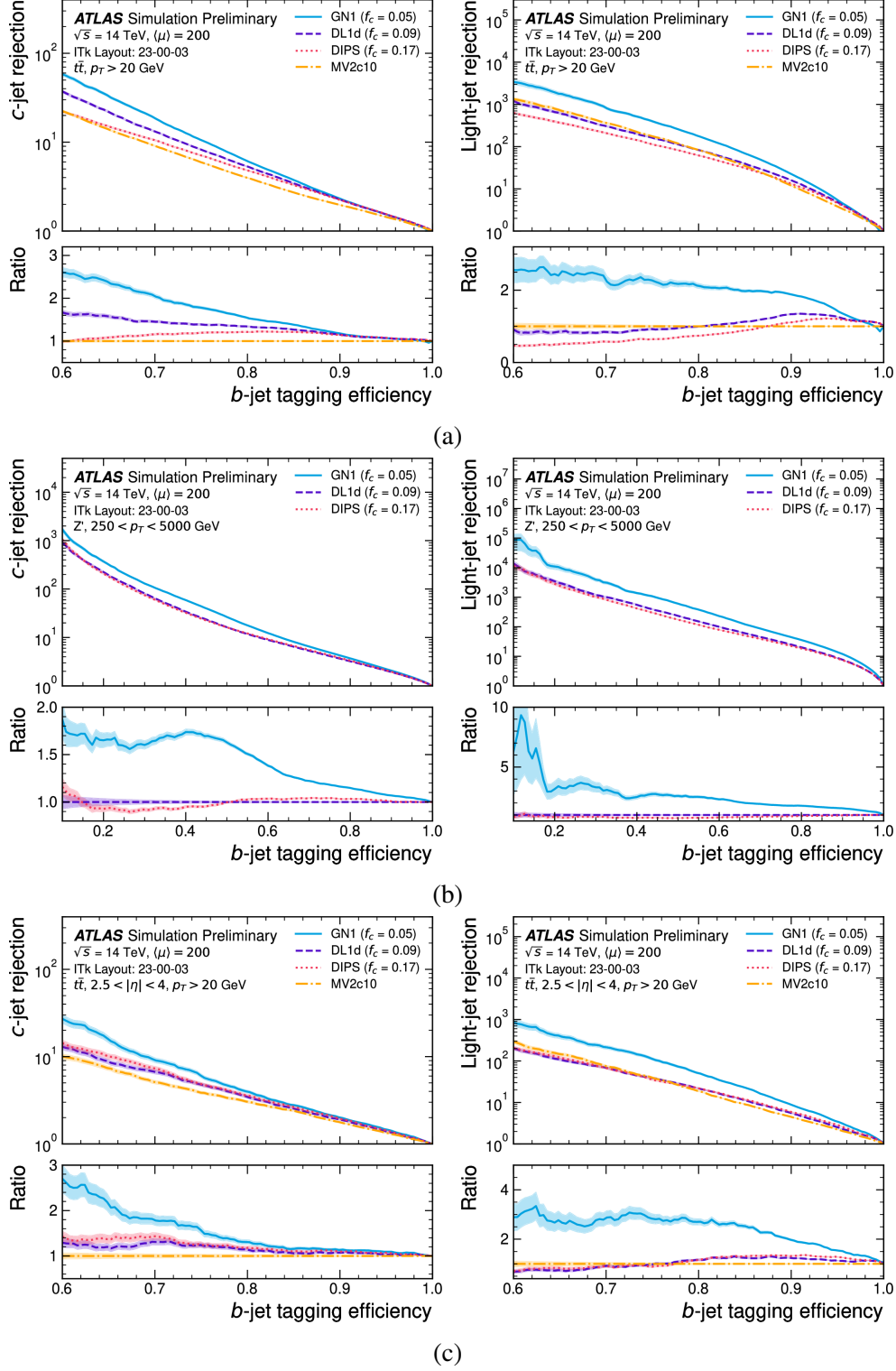
Figure 81: GN1, DL1d and the MV2 high-level tagger performances on the $t\bar{t}$ and Z' samples list in Table 13 [15].

GN1 Input Variables	
Jet Input	Description
P$_T$	Jet transverse momentum
η	Signed jet pseudorapidity

Track Input	Description
q/p	Track charge divided by momentum (measure of curvature)
d$_\eta$	Pseudorapidity of the track, relative to the jet η
d$_\phi$	Azimuthal angle of the track, relative to the jet ϕ
d$_0$	Closest distance from the track to the PV in the longitudinal plane
$z_0$sinθ	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
s(d$_0$)	Lifetime signed transverse IP significance
s($z_0$sinθ)	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nStripHits	Number of strip hits
nInnermostPixHits	Number of hits from the innermost pixel layer
nNextToInndermostPixHits	Number of hits from the next-to-innermost pixel layer
nInnermostPixShared	Number of shared hits from the innermost pixel layer
nInnermostPixSplit	Number of split hits from the innermost pixel layer
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nStripShared	Number of shared strip hits
nPixHoles	Number of pixel holes
nStripHoles	Number of strip holes

Table 24: Input features to the GN1 model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used [15].

APPENDIX B: Run2 Datasets and HLT Lepton Triggers for Anomaly Detection

All the data used in this analysis is originated from the ATLAS detector during the Run 2 period between the years 2015 to 2018. The dat was recorded during stable beam conditions while all relevant subdetectors were fully operational. The event candidates selected was done by either single-electron triggers or single-muon triggers which range in transverse momenta, transverse energy, and quality and isolation thresholds. The preselected data derive from the Good Run Lists (GRL) for 2015-2018 for release 21:

```
2015:  data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good
_25ns.xml
```

```
2016:  data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_
StandardGRL_All_Good_25ns.xml
```

```
2017:  data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All
_Good_25ns_Triggerno17e33prim.xml
```

```
2018:  data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All
_Good_25ns_Triggerno17e33prim.xml
```

The files process from these GRL lists come in a format called DAOD_STDM4 and required at least one lepton HLT trigger for either electron or muon HLTs. All saved objects within these files required at least a transverse momenta above 20 GeV for further preselection. The datasets in this

STDM4 derivation format used for this analysis are:

```
data15_13TeV.period[D-J].physics_Main.PhysCont.DAOD_STDM4.grp23_v01_p4238

data16_13TeV.period[A-G,I,K,L].physics_Main.PhysCont.DAOD_STDM4.grp23_v01_p4238

data17_13TeV.period[B-F,H,I,K].physics_Main.PhysCont.DAOD_STDM4.grp23_v01_p4238

data18_13TeV.period[B-D,F,I,K-M,O,Q].physics_Main.PhysCont.DAOD_STDM4.grp23_v01_p4238
```

Mini-trees were created from these dataset containers using the framework xAODAnaHelpers (AnalysisBase,21.2.177) and the Wjet framework that was developed in release 21.2.177, selecting events with leptons that pass the HLT triggers. The triggers in which dominate consistent efficiency over the data-taking period are the single lepton triggers with $P_T^l > 60$ GeV. The naming convention for these triggers follows `HLT_muNN[_isoInfo]`, where `NN` specifies the P_T threshold and the `isInfo` is the isolation requirement. Similarly for electron, the trigger naming convention is such `HLT_eNN_IDinfo[_lhInfo][_isoInfo]` where the [IDinfo] is the information on the identification requirement, `lhInfo` is an option requirement that enters the identification likelihood calculation and the `isoInfo` is the isolation requirement. The HLT triggers used in this analysis for both electron and muons are as following:

Muon Triggers

```
HLT_mu24_iloose || HLT_mu24ivarloose || HLT_24ivarmedium ||
HLT_mu26_ivarmedium || HLT_mu24imedium || HLT_mu26_imedium HLT_mu40 || HLT_mu50
```

Electron Triggers

```
HLT_e26_lhtight_nod0_ivarloose || HLT_e24_lhmedium_nod0_iverloose ||
HLT_e60_medium || HLT_e60_lhmedium_nod60
```

APPENDIX C: Monte Carlo Samples for the Background Hypothesis

The dominant source of background samples used for this analysis are $t\bar{t}$ and signle-top events along with W+jets simulated processes. Multi-jet processes are estimated using a loose lepton control region from data. The $t\bar{t}$ events are from the sample:

`mc16_13TeV.410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad.deriv.DAOD_STDM4.%p4237`

After requiring at least one lepton above 60 GeV, 25M events were left.

The single top samples used were:

`mc16_13TeV.410644.PowhegPythia8EvtGen_A14_singletop_%.DAOD_STDM4.%p4237`

`mc16_13TeV.410646.PowhegPythia8EvtGen_A14_Wt_DR_inclusive_top.%.DAOD_STDM4.%p4237`

`mc16_13TeV.410647.PowhegPythia8EvtGen_A14_Wt_DR_inclusive_antitop.%.DAOD_STDM4.%p4237`

`mc16_13TeV.410658.PhPy8EG_A14_tchan_BW50_lept_top.deriv.DAOD_STDM4.%p4237`

`mc16_13TeV.410659.PhPy8EG_A14_tchan_BW50_lept_antitop.deriv.DAOD_STDM4.%p4237`

These include three different processes, single-top, l+t, and W+t. All these are considered "single-top" processes within this analysis. The total amount of events from these datasets after requiring at least one lepton above 60 GeV is about 6M.

The W+jets samples are:

`mc16_13TeV.361100.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusenu.deriv.DAOD_STDM4.%p4237`

`mc16_13TeV.361103.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusenu.deriv.DAOD_STDM4.%p4237`

`mc16_13TeV.361101.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusmunu.deriv.DAOD_STDM4.%p4237`

`mc16_13TeV.361104.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusmunu.deriv.DAOD_STDM4.%p4237`

After requiring at least one lepton above 60 GeV, these samples contained about 4M events. There are 140M events from the data taken in Run2 within the defined signal region after the final selection described in section 6.2.4. The total amount of Monte Carlo simulated events after the final selection is about 14M, so about 10% of data. Once the loose lepton control region is added from the data, this gives sufficient amount of events to conclude a smooth falling background shape. These samples were processed using the full ATLAS detector simulation based on Geant4 and the pile-up conditions were matched per data-taking year.

APPENDIX D: RMM Event Examples

Figure 82 - 87 shows single events converted to RMM. The 9 invariant masses of interest are re-moved.
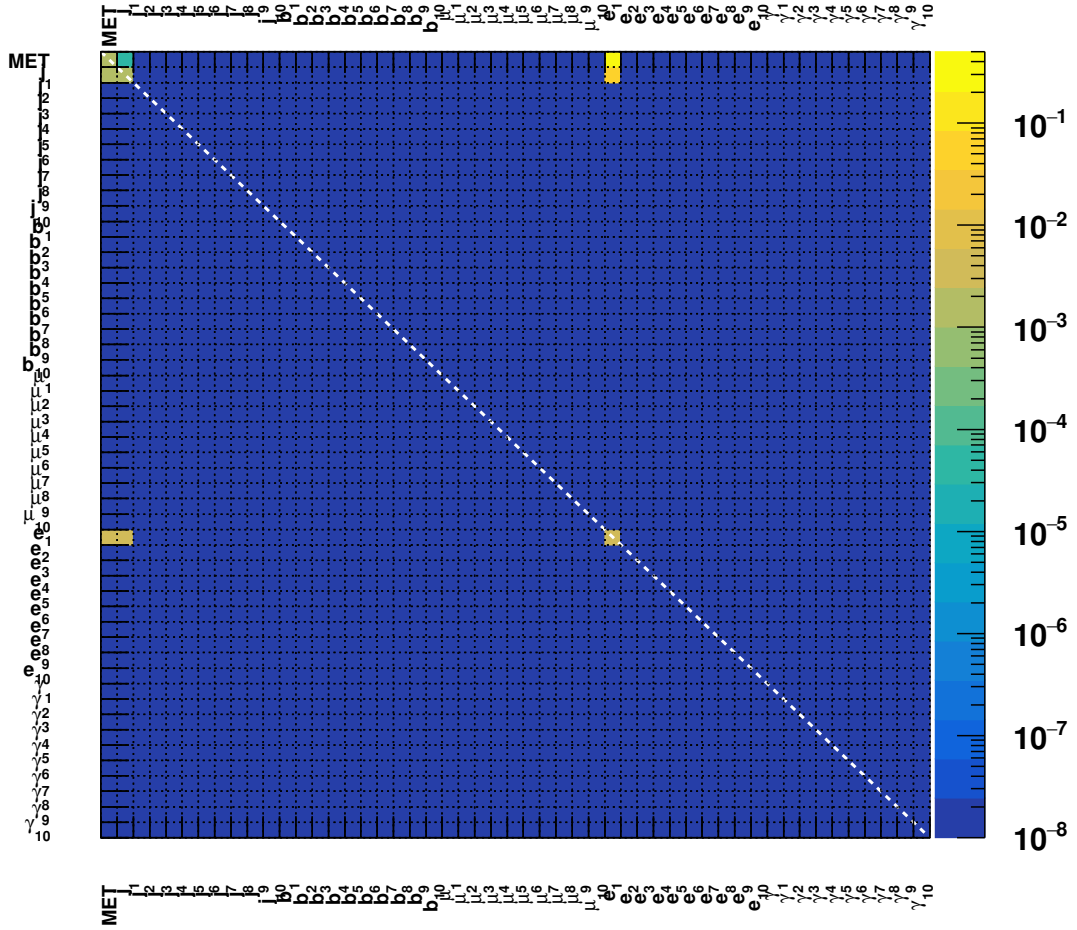
Figure 82: A typical data event (from 2016) shown as RMM. The event has one jet, one electron and some (small) MET.
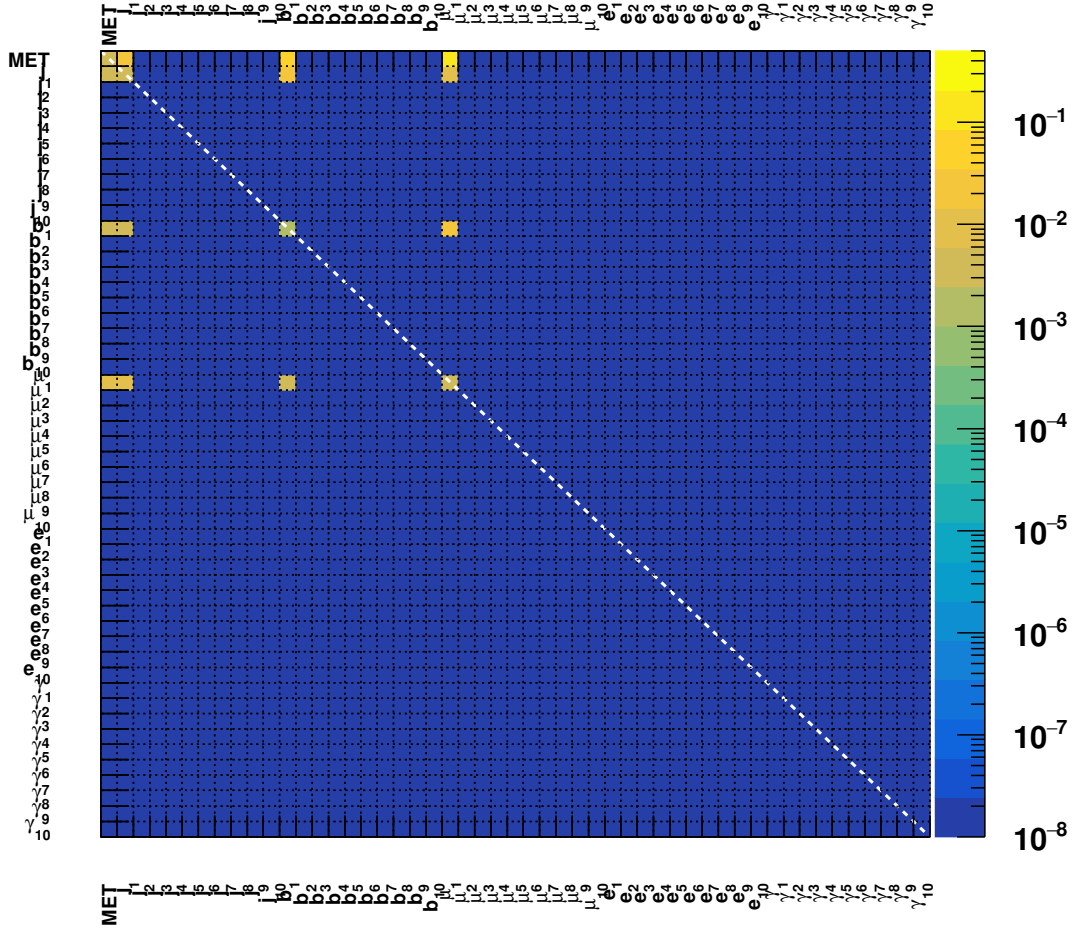
Figure 83: A typical $t\bar{t}$ event from a Monte Carlo simulation. The event has one jet, one **b**-jet, one muon and some MET.
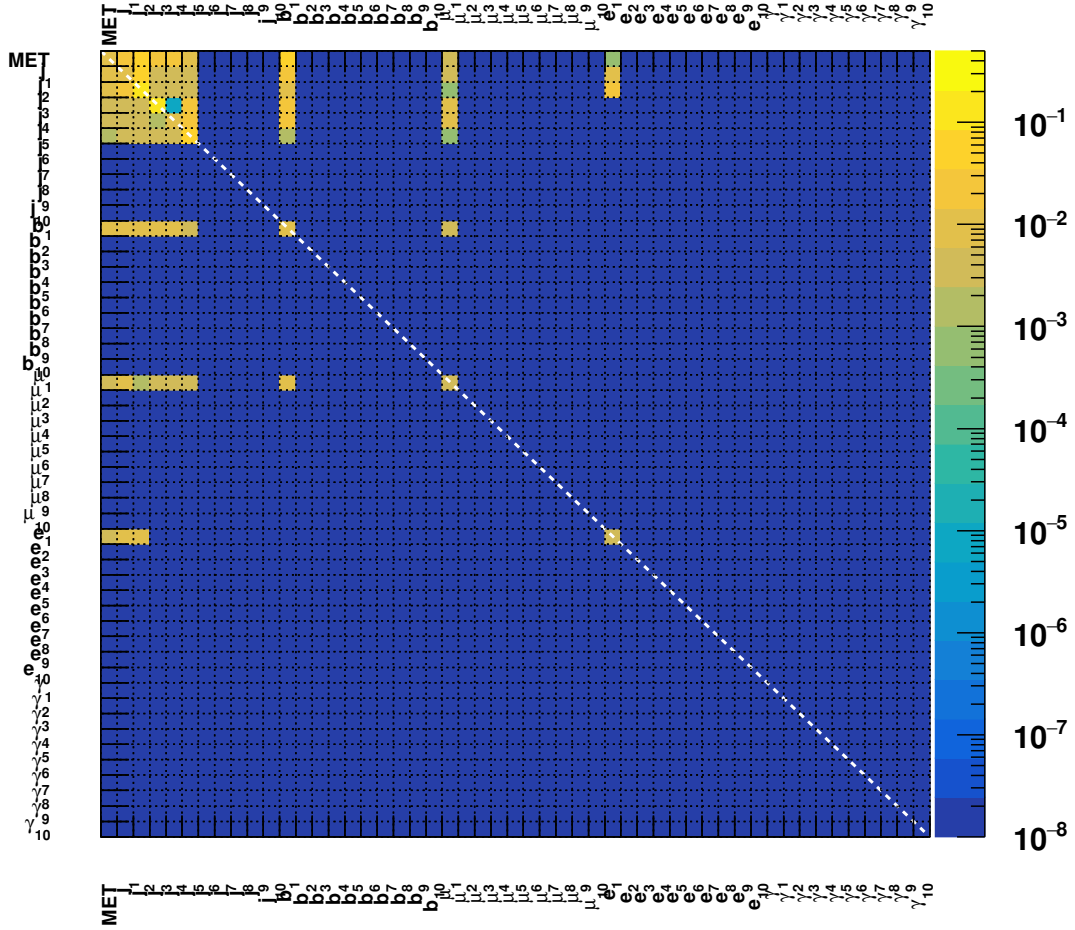
Figure 84: A typical the sequential standard model $W' \rightarrow WZ' \rightarrow l\nu q\bar{q}$ event with W' at 0.75 TeV and Z' at 0.5 TeV decaying to 2 jets, with the leptonic decay of W. The event has multiple jets, leptons and some MET.
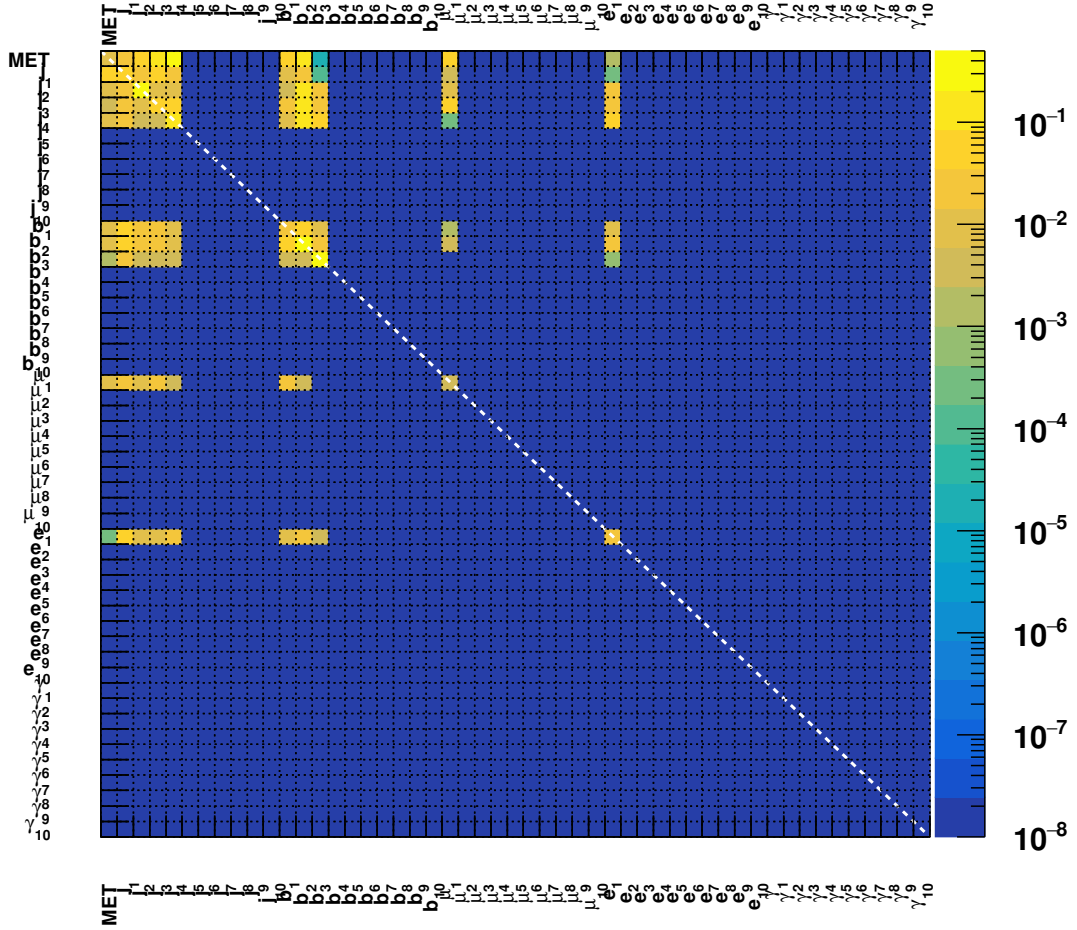
Figure 85: A typical event for the charged Higgs production in association with a top quark, tbH^+. The mass of H^+ is set to 2 TeV. decaying to 2 jets, with the leptonic decay of W. The event features many jets, $b-$jets, leptons and some MET.
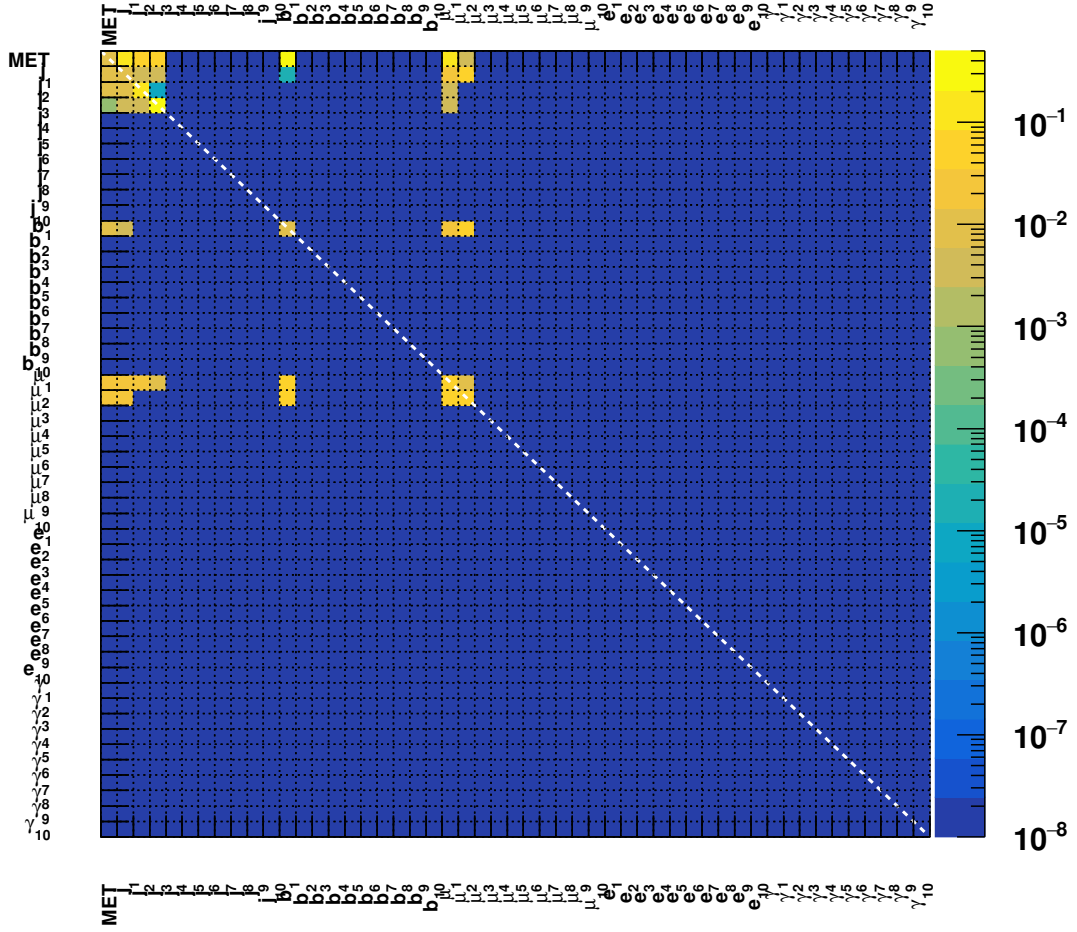
Figure 86: A typical event for a composite lepton E from a decay of massive Z' with various Z' mass hypotheses. The mass of Z' is set to 3 TeV. The event features jets, $b-$jets, 2 muons and some MET.
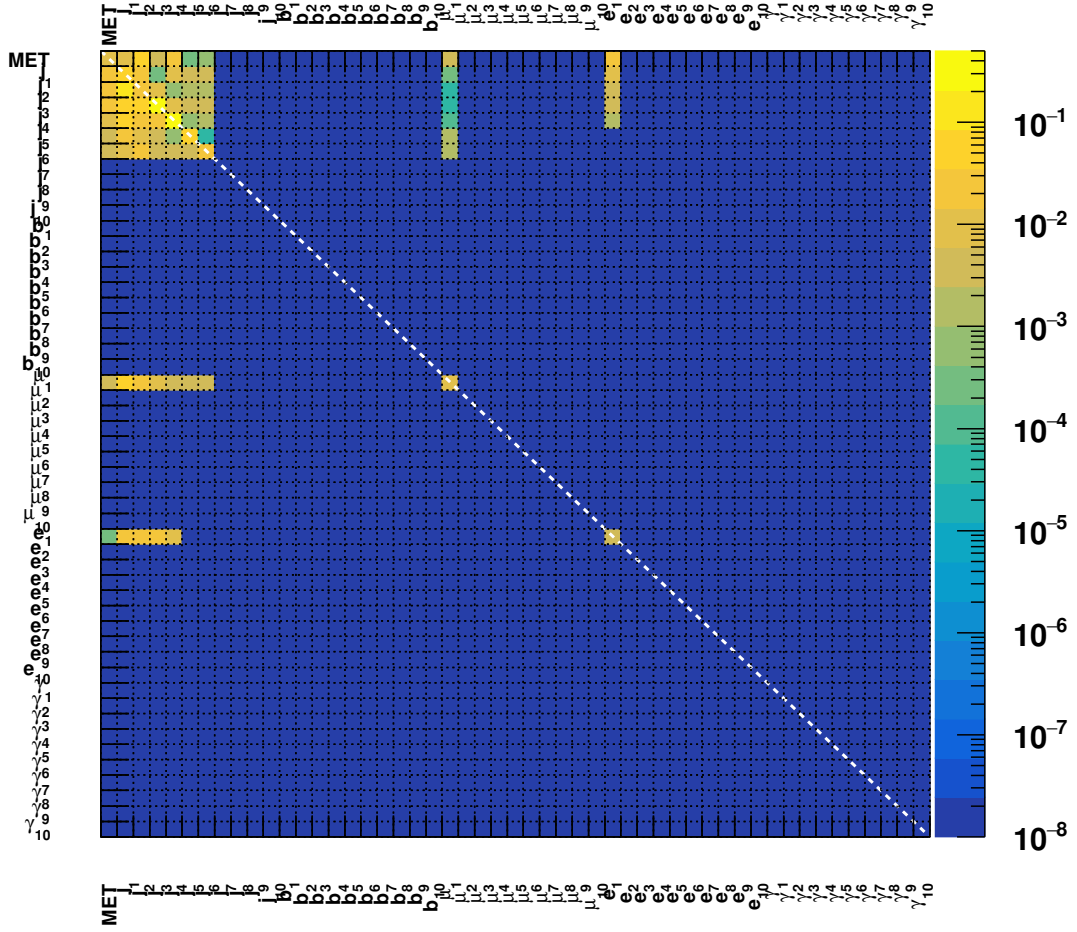
Figure 87: A typical event for a Kaluza–Klein (KK) gauge boson, W_{kk}, with a SM W boson and a radion. The mass of W_{kk} is set to 4 TeV. The event features many jets, $b-$jets, leptons and some MET.

APPENDIX E: Autoencoder Topology Studies

Ten different topologies were studied for this analysis along with different types of AEs such as variational AE and the convolutional AE. The optimization metrics that were focused on was the separation power between SM and BSM events, along with loss value range. Having a larger range of loss values indicates the every small nuance within the data representation plays a role in data separation, which is key for anomaly detection. The separation corresponded with the number of neurons within the dense layers of the encoder and decoder. The nominal architecture may have been increased but was capped due to available computational power.

Hand-made anomalies were created using the W' / Z' BSM events with the mass of 4 TeV. Anomaly 1 is when all jets beyond the sub-leading jet were set to he photons and anomaly 2 is when all the jets beyond the sub-leading jet are set to b-jets. High multiplicity of photons and b-jets are considered rare within the SM and are expected to be scored as anomalous. Figure 88 shows each of these three cases.

(a) SSM model   (b) Anomaly 1   (c) Anomaly 2
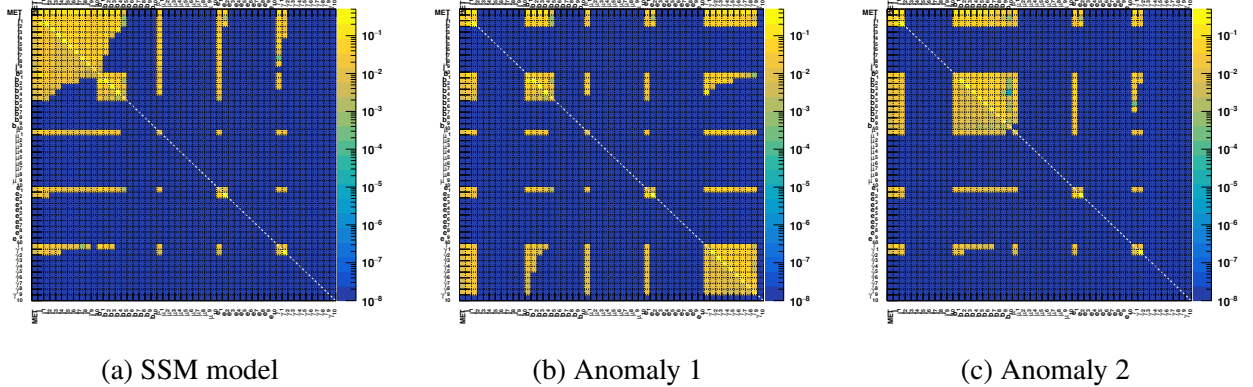
Figure 88: An example RMM matrix of a random event from SSM (a), from which jets beyond the sub-leading jet are set to be photons (b) or *b*-jets (c).

The nominal AE architecture as seen in Table 20 is "800-400-200-400-800" is used to predict the loss values for these samples in Figure 89. The loss value take for scoring anomalies is the Log-

181

arithm of the loss value in order to obtain human friendly values while also having a larger distribution. Figure X shows the nominal architecture along with three others ("400-200-100-200-400", "200-100-50-100-200", and "100-50-25-50-100") for comparison. The bottom two loss plots also show the two types of AEs (VAE and CVAE). As seen from these figures, the nominal architecture shows the best separation between SSM events ("Before anomaly") and the hand-made anomalies as discussed above.
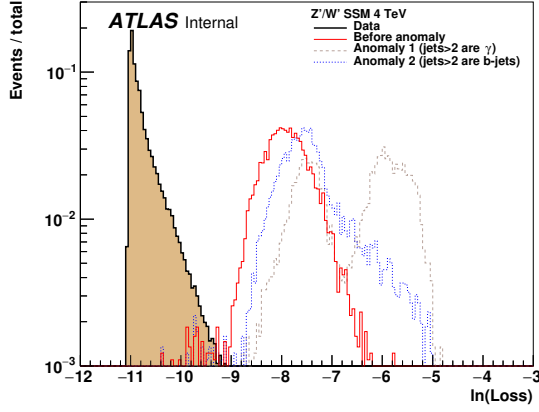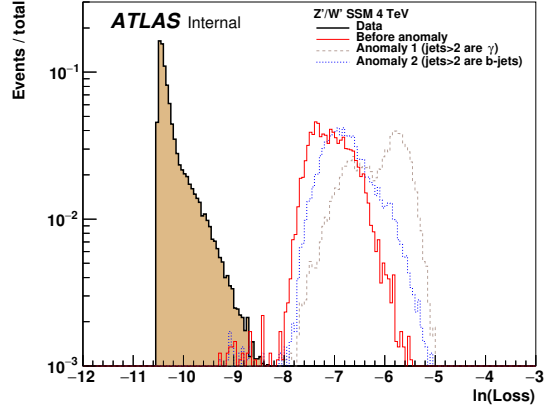
(a) 800-* topology   (b) 200-* topology
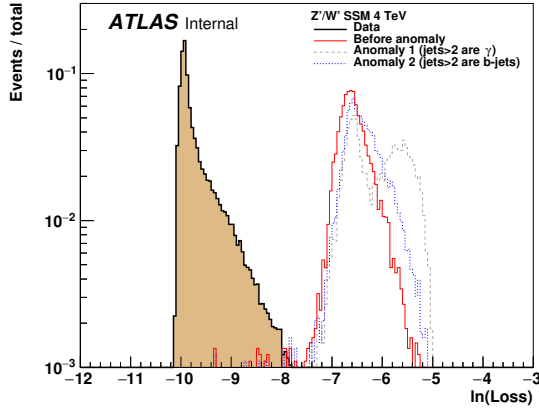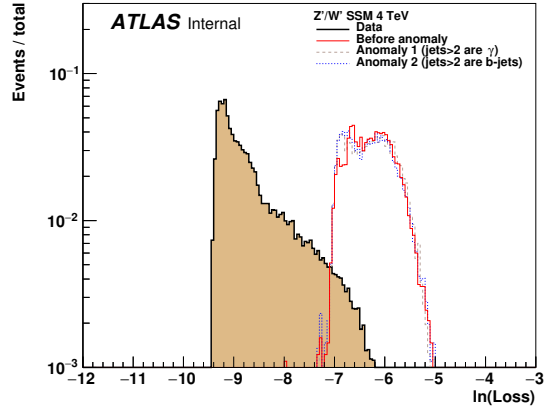
(c) 100-* topology   (d) 20-* topology
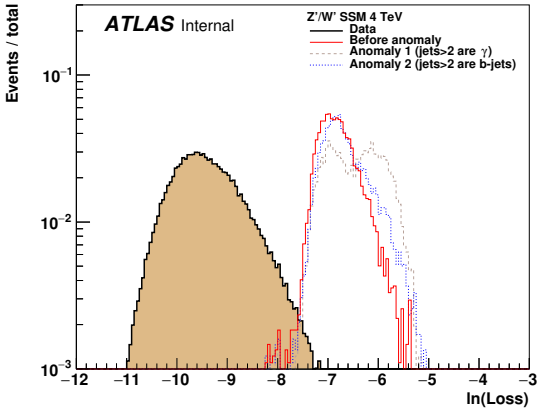
(e) Variational AE topology   (f) Convolutional VAE topology
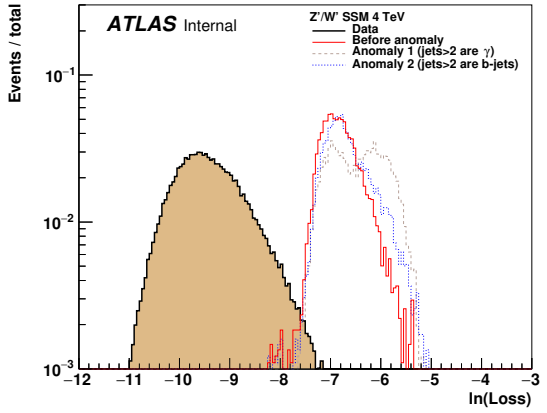
Figure 89: The loss distributions for the original SSM with 4 TeV Z' and for anomaly 1 and 2.

APPENDIX F: Alternative AE Models for Systematics

As discussed in Section 6.5, three models were training using the optimized architecture after 50 trainings were conducted. The nominal refers to the mean loss model, the "up" and "down" models correspond to the slightly higher anad lower loss values. The performances of these three models are evaluated on BSM signals in Figure 90. The outlier regions that are used will have low signal efficiency (low x on these plots). The performances are consistent.
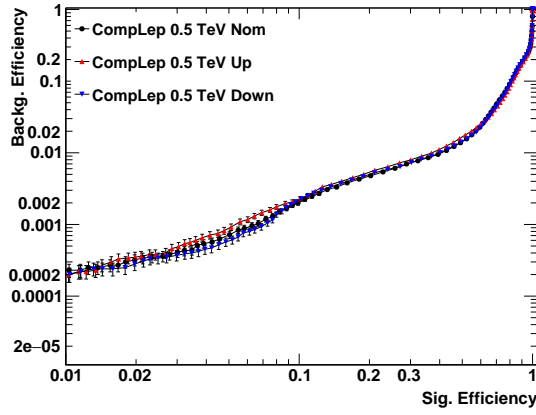
(a) Composite leptons low mass
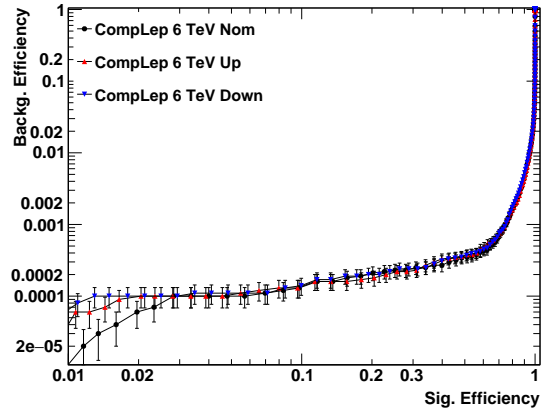
(b) Composite leptons high mass
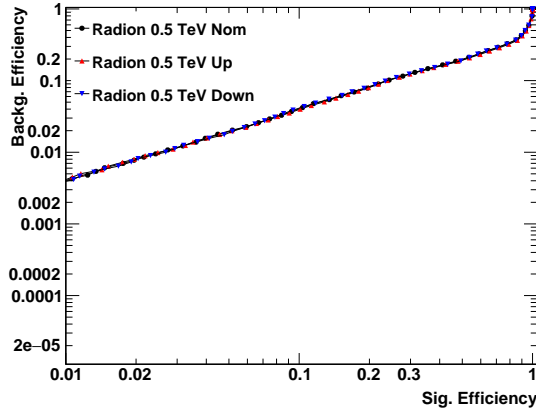
(c) Radion low mass
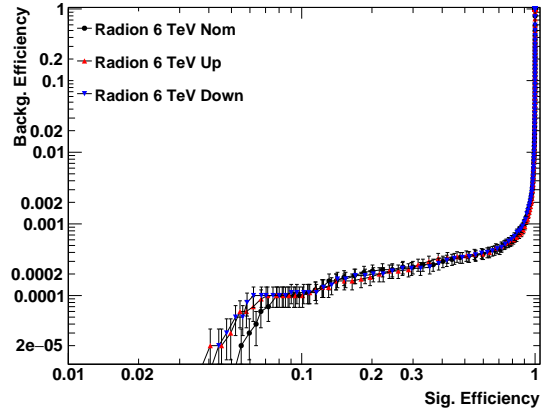
(d) Radion high mass
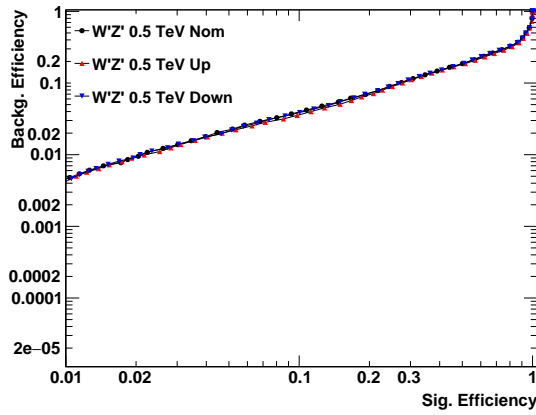
(e) W'/Z' low mass
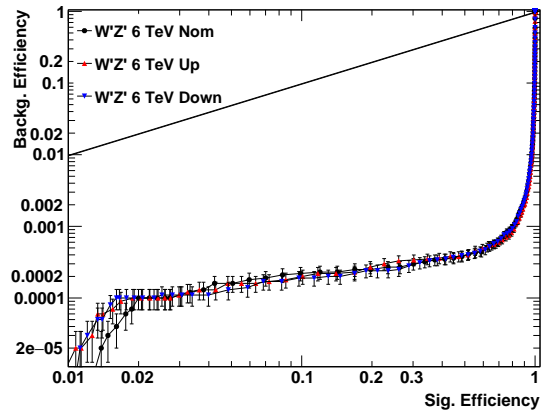
(f) W'/Z' high mass

Figure 90: Background efficiency vs signal efficiency of various BSM models under different mass hypotheses, using the nominal and alternative AE models. Note there are some artificial lines due to plotting issues.

APPENDIX G: S/B Improvement Example

There are many factors that determine the how well the S/B increases after an anomaly score cut, such as topology of the final states, the mass of heavy particles, the type of di-object invariant mass, the chosen AR working point, etc. It is expected that the more "exotic" the BSM model is, the more it deviates from the SM background. Therefore, it's expected that the heavier the BSM resonance is, the higher the anomaly score will be given, resulting in a large S/B calculation.

The following is an example using the radion BSM model due to its rather complex decay topology in two-body (even three-body) mass. The radion particle decays from a Kaluza-Klein boson Wkk into a radion denoted by φ as seen in:

$$Wkk \rightarrow W + \varphi \rightarrow l\nu + gg$$

where the events should have MET as there is a neutrino due to the W decay. The invariant mass that's focused on within this example is the jet+electron (m_{je}) since the final state is a semi-leptonic decay. The mass resonance used for the Wkk boson is 2 TeV decaying to a $\varphi = 500$ GeV. Figure X shows the signal event yields before and after the 10 pb AR cut while also showing the MC background yields that include the loose electron control region (LE-CR) which is a data-driven multi-jet estimation (discussed in Section X). Table 25 shows the numerical values for these yields after the 10 pb AR. The S/B is increased by $\tilde{5}00\%$. The discovery sensitivity [34] is calculated using the Eq. .1.

$$Z_A = \sqrt{2\left((s+b)\ln\left(1 + \frac{s}{b}\right) - s\right)} \tag{.1}$$

is improved by $Z_A = {\sim}80\%$. After the 1 pb cut, the S/B is incrased by ${\sim}1100\%$ and $Z_A = {\sim}12\%$.
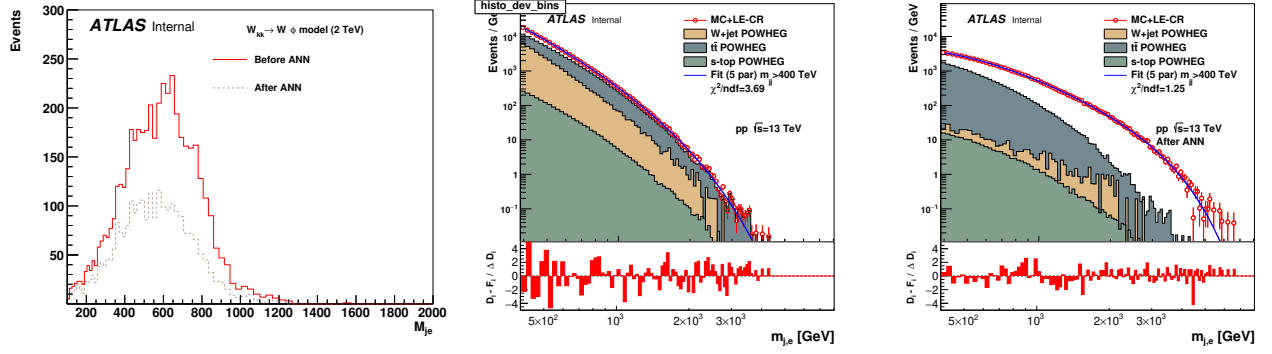
(a) Radion before and after AR cut

(b) Before AR cut

(c) After AR cut

Figure 91: (a) 500 GeV radion model signal yields before and after the 10 pb AR cut is applied. (b)(c) Comparison background events for the jet+electron invarint mass before and after the 10 pb AR cut.

	Before cut	10 pb	10 pb / Before	1 pb	1 pb / Before	0.1 pb	0.1 pb / Before
Radion 2 TeV	3356	1799	0.53	338	0.101	65	0.0193
Bkg	8999316	790230	0.087	72206	0.0080	7280	0.00081
S/B	0.00037	0.0022	6.10	0.0047	12.5	0.0089	24.1
Z_A	1.11	2.02	**1.80**	1.25	**1.12**	0.76	**0.68**

Table 25: Sensitivity gain of the radion model with Wkk set to 2 TeV before and after applying the 10 pb and 1 pb AR cut. Event yields counted in the 400-800 GeV range for the invariant mass m_{je}.

APPENDIX H: Statistical Fit Function Studies for 1pb AR

Similar to Section 6.6.3, this appendix shows the studies conducted for the 1 pb AR. Since the 5p function showed to be the reasonable out of the three tested functions int the 10 pb AR region, only the 4p and 5p fit functions are studied for the 1 pb and 0.1 pb AR. These plots show the functions in the MC+LE-CR. This region has noticable less statistics than the 10 pb region, therefore it's prone to larger statistical fluctuations within the fit functions, especially when requiring a b-jet or photon. Due to the minimal amount of statistics, a reasonable fit function that could be used for all nine invariant masses was unable to be found. A summary for the 1 pb statistical tests can be found in Table X. The following figures show the fits of the p4 and p5 functions along with their pulls for all nine invariant masses. The fit studies for the 0.1 pb region can be found in Appendix X.

- Figure 92 1pb m_{jj}
- Figure 95 1pb m_{je}
- Figure 98 1pb m_{be}

- Figure 93 1pb m_{jb}
- Figure 96 1pb $m_{j\mu}$
- Figure 96 1pb $m_{b\mu}$

- Figure 94 1pb m_{bb}
- Figure 97 1pb $m_{j\gamma}$
- Figure 100 1pb $m_{b\gamma}$

Mass	region	p	Fit χ^2	Pull μ	$\Delta\mu$	σ	$\Delta\sigma$	Gaus χ^2	KS	Shapiro
m_{jj}	1 pb	4	1.022132	0.105702	0.167736	1.053082	0.153271	0.839537	0.308644	0.978067
m_{jb}	1 pb	4	0.922279	-0.074200	0.134272	0.997789	0.127777	1.371910	0.686471	0.985667
m_{bb}	1 pb	4	0.709970	0.156393	0.119468	0.883445	0.107321	0.206520	0.162686	0.981281
m_{je}	1 pb	4	1.111753	0.322995	0.114167	0.931633	0.147120	0.806860	0.304359	0.951957
$m_{j\mu}$	1 pb	4	1.555536	0.876259	0.138686	0.854421	0.265782	1.496423	0.000015	0.910287
$m_{j\gamma}$	1 pb	4	1.068962	0.312716	0.180434	1.165488	0.362410	1.419774	0.493023	0.973990
m_{be}	1 pb	4	2.007666	0.964171	0.168553	0.979930	0.189254	1.267569	0.000000	0.898834
$m_{b\mu}$	1 pb	4	1.281002	0.665929	0.129382	0.751723	0.227398	1.066046	0.004606	0.966401
$m_{b\gamma}$	1 pb	4	1.282447	0.404946	0.129898	0.878308	0.168753	0.566677	0.050555	0.957157
m_{jj}	1 pb	5	0.996736	0.206397	0.123558	1.000077	0.117074	0.624787	0.407666	0.978652
m_{jb}	1 pb	5	0.931768	0.026414	0.129550	1.018507	0.111694	0.889204	0.852813	0.988362
m_{bb}	1 pb	5	0.740004	0.254771	0.107926	0.790410	0.099640	0.741691	0.128867	0.979232
m_{je}	1 pb	5	1.079768	0.378016	0.095334	0.790995	0.095697	0.954058	0.223397	0.939962
$m_{j\mu}$	1 pb	5	1.287571	0.365692	0.193126	1.091768	0.237806	1.412079	0.001635	0.928943
$m_{j\gamma}$	1 pb	5	1.081601	0.520053	0.206680	1.157502	0.409198	1.705933	0.297592	0.972306
m_{be}	1 pb	5	1.777508	0.801023	0.280586	1.518162	0.305243	1.128937	0.000001	0.925694
$m_{b\mu}$	1 pb	5	1.305168	0.650556	0.144944	0.740383	0.263902	1.202923	0.001945	0.965139
$m_{b\gamma}$	1 pb	5	1.252325	0.463297	0.137670	0.828874	0.276603	1.000151	0.070159	0.968573

Table 26: Statistical quantities for SM MC+LE-CR fit for the 1 pb AR.

(a) m$_{jj}$ using MC+LE-CR, p4

(b) pulls of m$_{jj}$ in p4

(c) m$_{jj}$ using MC+LE-CR, p5

(d) pulls of m$_{jj}$ in p5

Figure 92: The m$_{jj}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

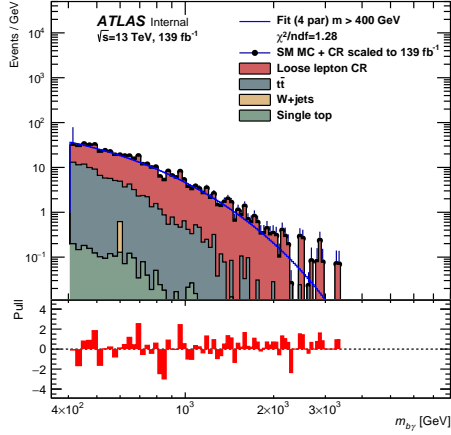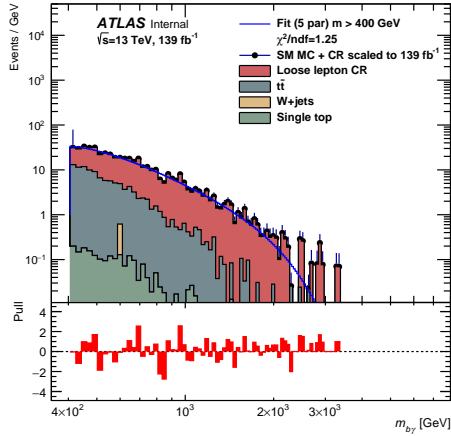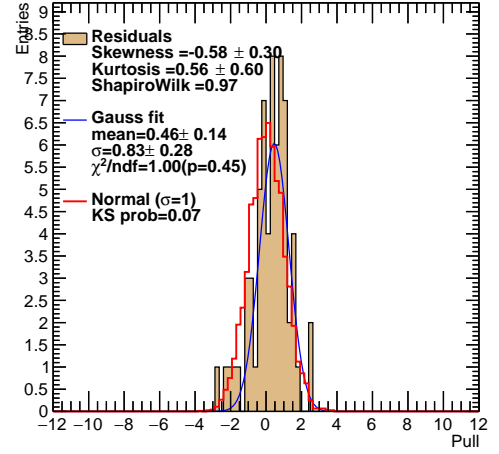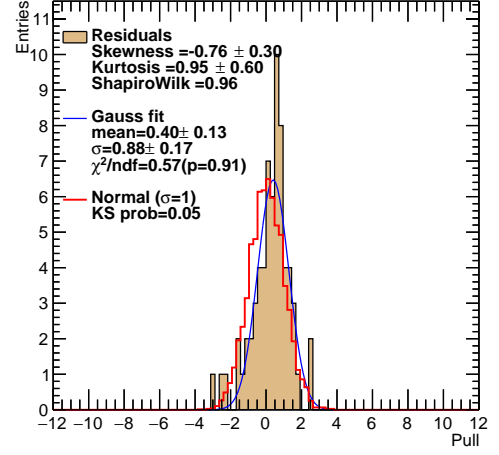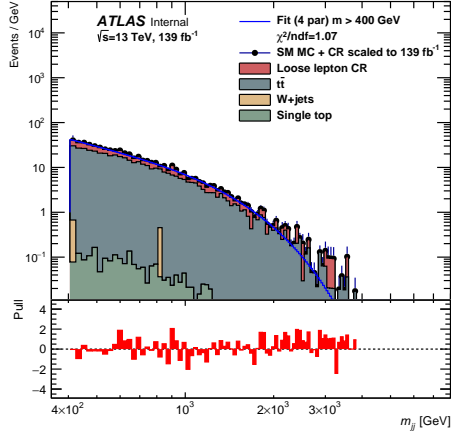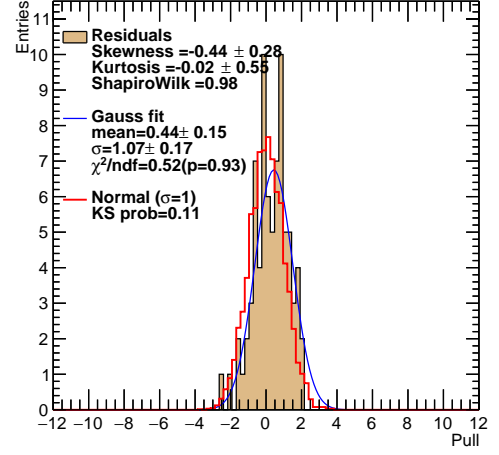(a) m$_{jb}$ using MC+LE-CR, p4

(b) pulls of m$_{jb}$ in p4

(c) m$_{jb}$ using MC+LE-CR, p5

(d) pulls of m$_{jb}$ in p5

Figure 93: The m$_{jb}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) m$_{bb}$ using MC+LE-CR, p4

(b) pulls of m$_{bb}$ in p4

(c) m$_{bb}$ using MC+LE-CR, p5

(d) pulls of m$_{bb}$ in p5

Figure 94: The m$_{bb}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.
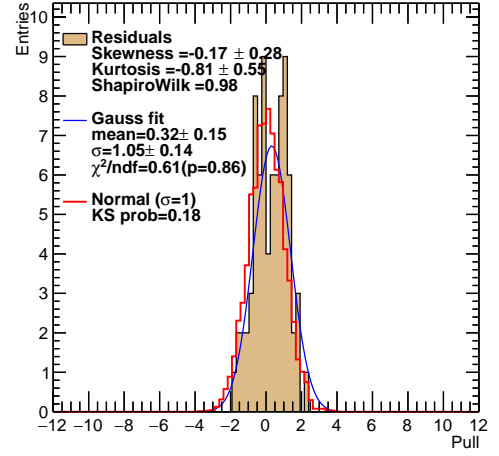
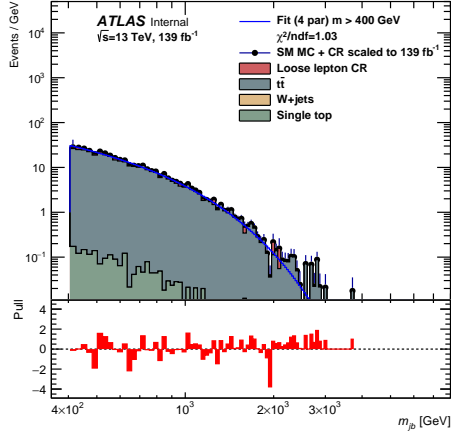(a) m$_{je}$ using MC+LE-CR, p4

(b) pulls of m$_{je}$ in p4

(c) m$_{je}$ using MC+LE-CR, p5

(d) pulls of m$_{je}$ in p5

Figure 95: The m$_{je}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) m$_{j\mu}$ using MC+LE-CR, p4

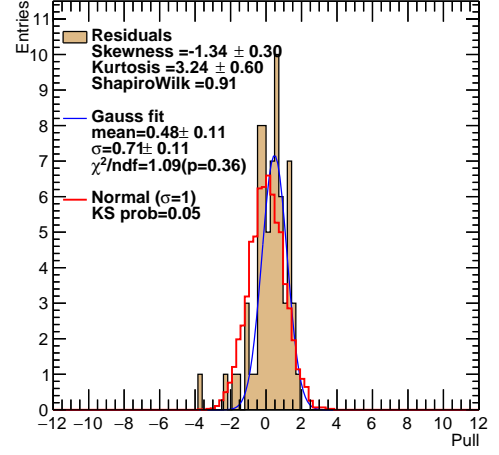(b) pulls of m$_{j\mu}$ in p4

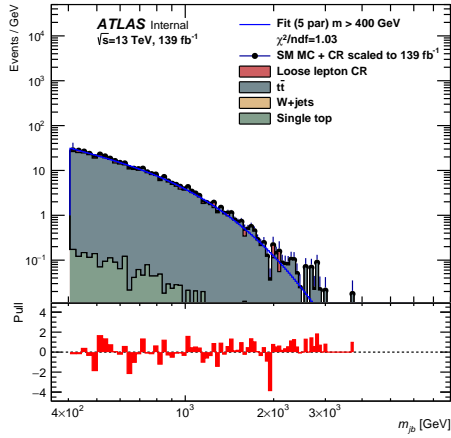(c) m$_{j\mu}$ using MC+LE-CR, p5

(d) pulls of m$_{j\mu}$ in p5

Figure 96: The m$_{j\mu}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.
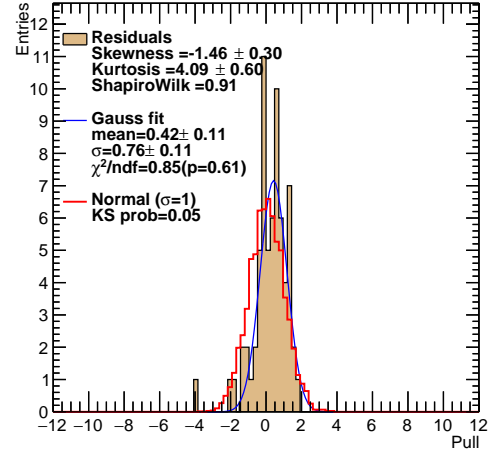
(a) $m_{j\gamma}$ using MC+LE-CR, p4

(b) pulls of $m_{j\gamma}$ in p4

(c) $m_{j\gamma}$ using MC+LE-CR, p5

(d) pulls of $m_{j\gamma}$ in p5

Figure 97: The $m_{j\gamma}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) m_{be} using MC+LE-CR, p4
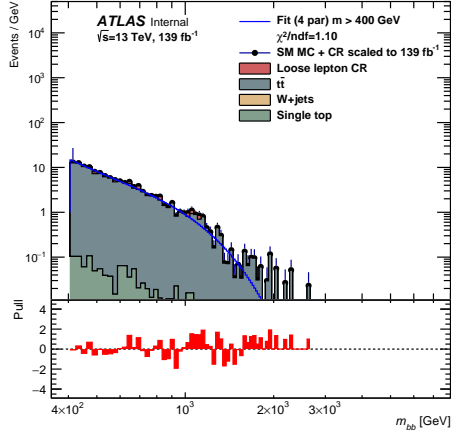
(b) pulls of m_{be} in p4
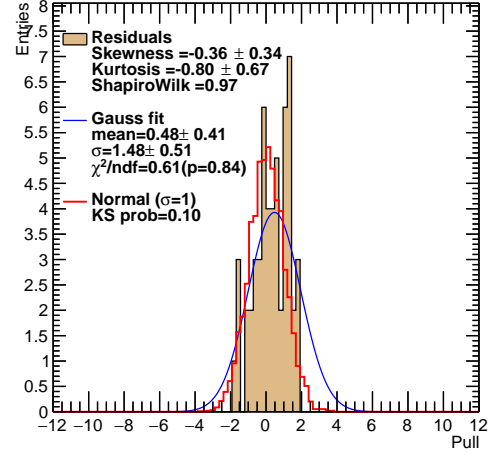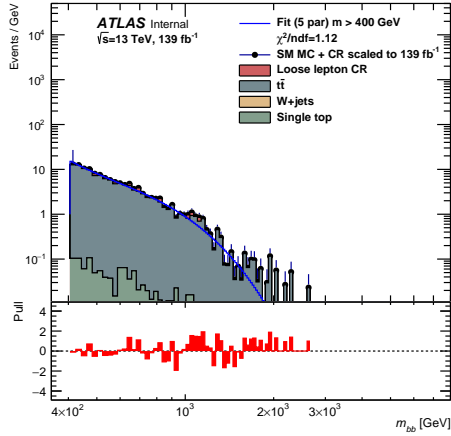
(c) m_{be} using MC+LE-CR, p5

(d) pulls of m_{be} in p5

Figure 98: The m_{be} invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) $m_{b\mu}$ using MC+LE-CR, p4

(b) pulls of $m_{b\mu}$ in p4

(c) $m_{b\mu}$ using MC+LE-CR, p5

(d) pulls of $m_{b\mu}$ in p5

Figure 99: The $m_{b\mu}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.
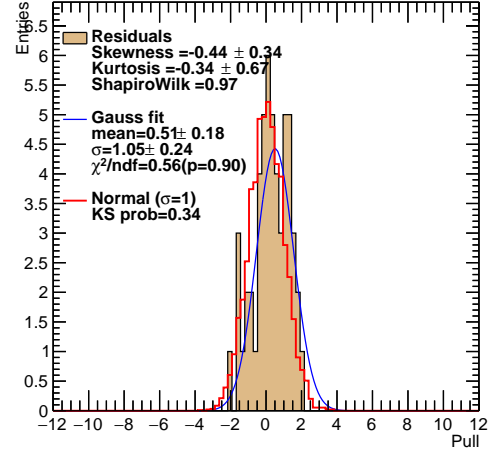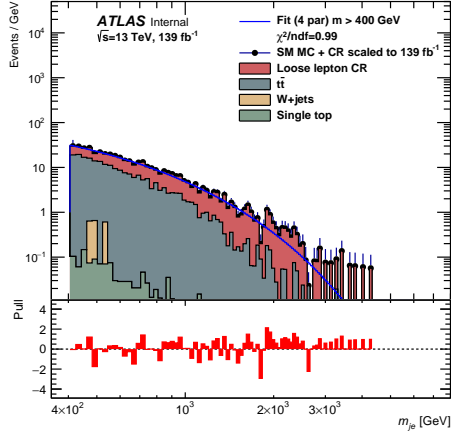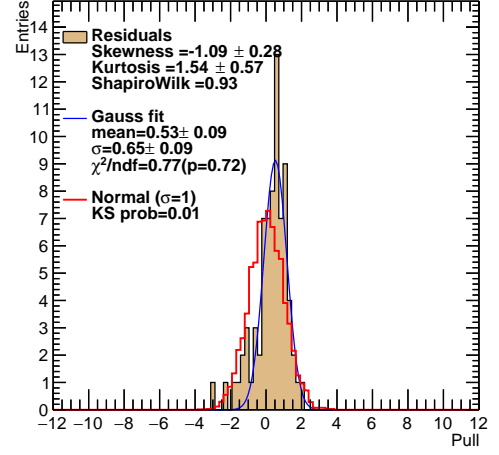
(a) $m_{b\gamma}$ using MC+LE-CR, p4

(b) pulls of $m_{b\gamma}$ in p4

(c) $m_{j\gamma}$ using MC+LE-CR, p5

(d) pulls of $m_{b\gamma}$ in p5

Figure 100: The $m_{b\gamma}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

APPENDIX I: Statistical Fit Function Studies for 0.1pb AR

Just as in Appendix APPENDIX H:, only the p4 and p5 fit functions are studied. The results can be seen in Table X and the figures show the MC+LE-CR along with their pulls. This region only has 14K events, therefore finding a reasonable fit function was inconclusive, just as in the 1 pb AR.

- Figure 101 0.1pb m_{jj}

- Figure 102 0.1pb m_{jb}

- Figure 103 0.1pb m_{bb}

- Figure 104 0.1pb m_{je}

- Figure 105 0.1pb $m_{j\mu}$

- Figure 106 0.1pb $m_{j\gamma}$

- Figure 107 0.1pb m_{be}

- Figure 105 0.1pb $m_{b\mu}$

- Figure 109 0.1pb $m_{b\gamma}$

Mass	region	p	Fit χ^2	Pull μ	$\Delta\mu$	σ	$\Delta\sigma$	Gaus χ^2	KS	Shapiro
m_{jj}	0.1 pb	4	1.066981	0.439052	0.154240	1.066725	0.168530	0.520639	0.114105	0.982610
m_{jb}	0.1 pb	4	1.030326	0.478555	0.107806	0.706963	0.108466	1.092364	0.054007	0.913377
m_{bb}	0.1 pb	4	1.104344	0.476400	0.405565	1.478379	0.507329	0.607558	0.104235	0.967501
m_{je}	0.1 pb	4	0.988952	0.527118	0.087665	0.647729	0.091518	0.773745	0.008889	0.934337
$m_{j\mu}$	0.1 pb	4	1.247852	0.688118	0.120311	0.710719	0.131556	0.958169	0.012524	0.894489
$m_{j\gamma}$	0.1 pb	4	0.766672	0.325899	0.104514	0.712873	0.103878	0.994939	0.093205	0.958485
m_{be}	0.1 pb	4	0.689806	0.222383	0.179162	0.911493	0.183285	0.341040	0.355303	0.965579
$m_{b\mu}$	0.1 pb	4	0.994301	0.577633	1.529995	2.193773	1.597088	1.426726	0.091036	0.934948
$m_{b\gamma}$	0.1 pb	4	0.532739	0.123343	0.153080	0.761905	0.157570	0.373211	0.509120	0.982694
m_{jj}	0.1 pb	5	0.982026	0.323078	0.147239	1.053884	0.137738	0.611454	0.178563	0.984839
m_{jb}	0.1 pb	5	1.032755	0.424803	0.110022	0.755037	0.111329	0.853758	0.046379	0.911716
m_{bb}	0.1 pb	5	1.115241	0.510098	0.184014	1.048742	0.241893	0.562814	0.337324	0.970615
m_{je}	0.1 pb	5	0.996203	0.232937	0.133625	0.900851	0.117146	1.034730	0.002092	0.934982
$m_{j\mu}$	0.1 pb	5	1.150431	0.443298	0.176520	0.758198	0.239118	0.846013	0.017779	0.889262
$m_{j\gamma}$	0.1 pb	5	0.764039	0.158883	0.118088	0.817065	0.101391	0.592739	0.412045	0.969030
m_{be}	0.1 pb	5	0.708264	0.186595	0.214779	0.993727	0.222814	0.713869	0.355303	0.965863
$m_{b\mu}$	0.1 pb	5	0.985081	0.139081	0.285717	1.249059	0.346504	0.645550	0.723507	0.950681
$m_{b\gamma}$	0.1 pb	5	0.619378	0.147059	0.163189	0.805034	0.202471	0.437407	0.329640	0.988531

Table 27: Statistical quantities for SM MC+LE-CR fit for the 1 pb AR.

(a) m$_{jj}$ using MC+LE-CR, p4

(b) pulls of m$_{jj}$ in p4

(c) m$_{jj}$ using MC+LE-CR, p5

(d) pulls of m$_{jj}$ in p5

Figure 101: The m$_{jj}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.
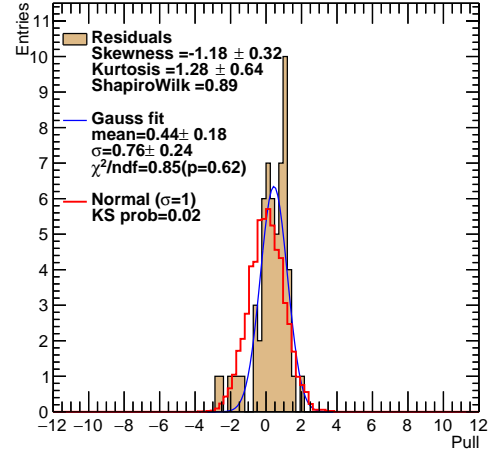
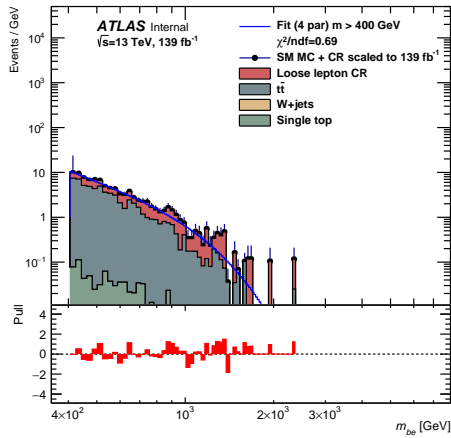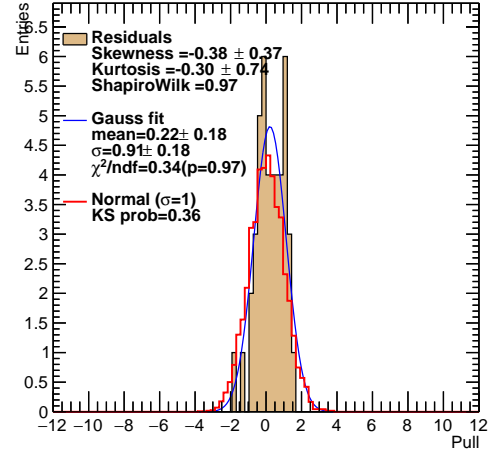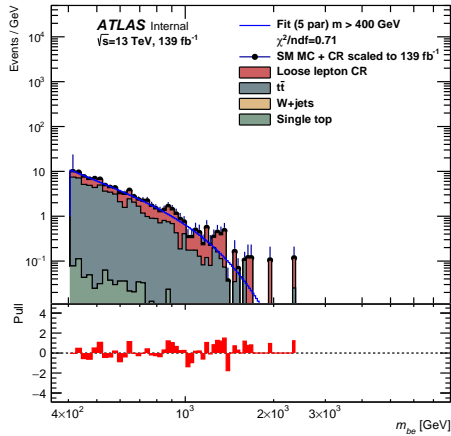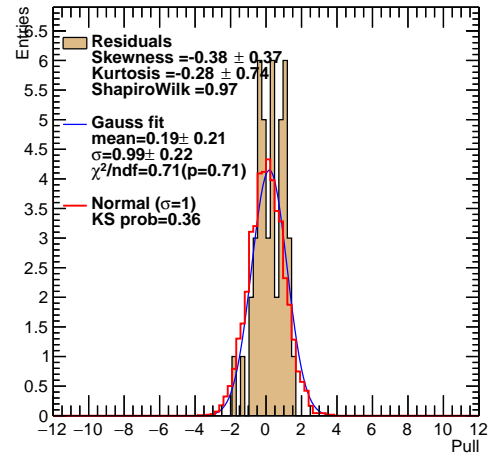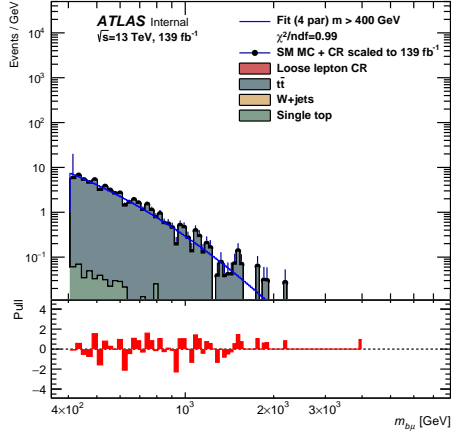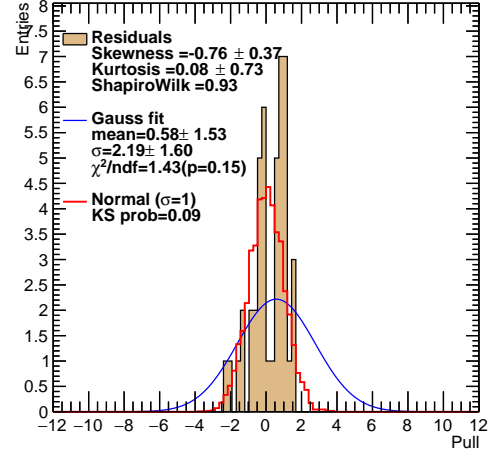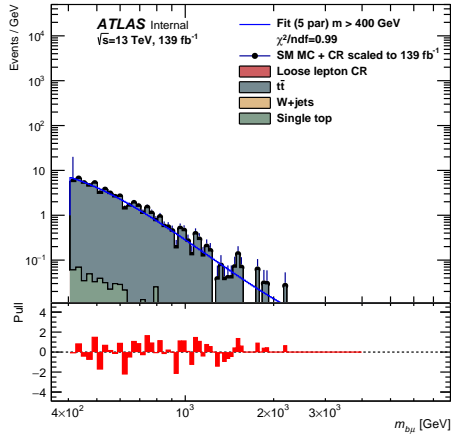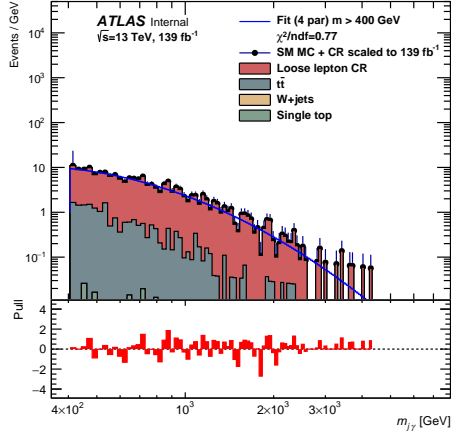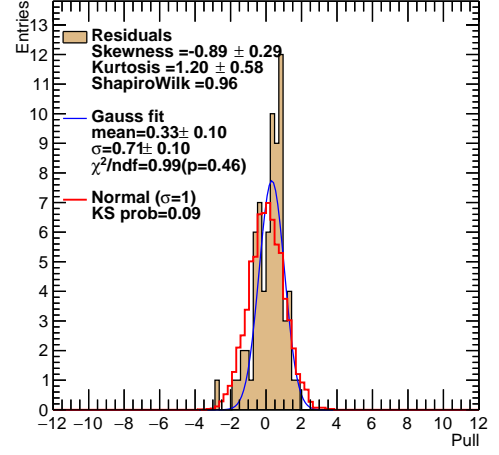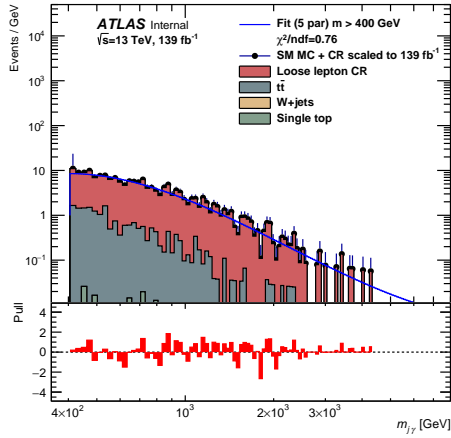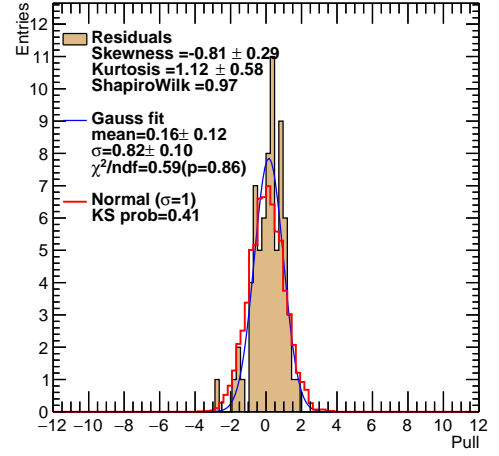(a) m$_{jb}$ using MC+LE-CR, p4

(b) pulls of m$_{jb}$ in p4

(c) m$_{jb}$ using MC+LE-CR, p5

(d) pulls of m$_{jb}$ in p5

Figure 102: The m$_{jb}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.
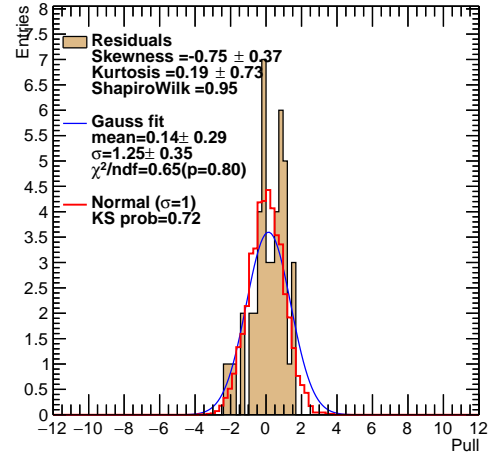
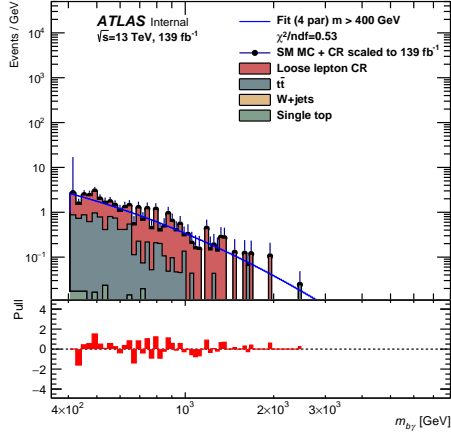(a) m$_{bb}$ using MC+LE-CR, p4

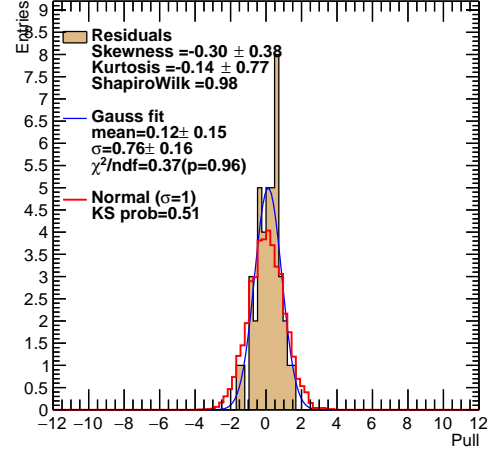(b) pulls of m$_{bb}$ in p4

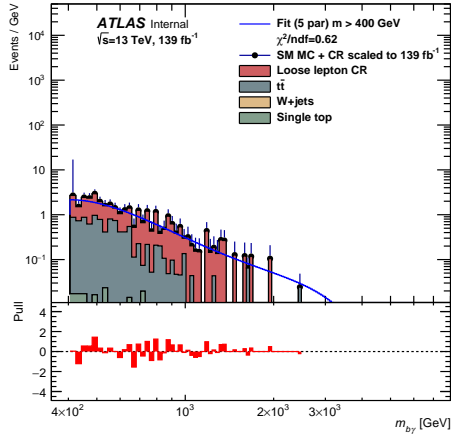(c) m$_{bb}$ using MC+LE-CR, p5

(d) pulls of m$_{bb}$ in p5

Figure 103: The m$_{bb}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) m_{je} using MC+LE-CR, p4
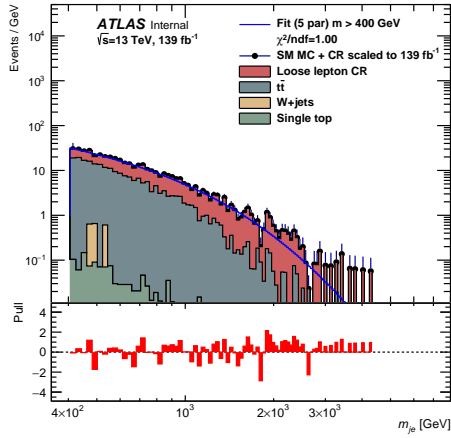
(b) pulls of m_{je} in p4

(c) m_{je} using MC+LE-CR, p5
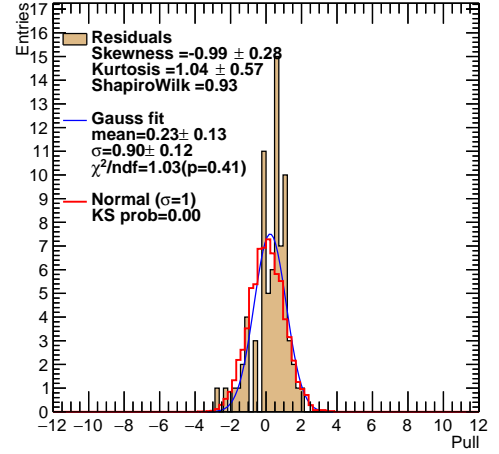
(d) pulls of m_{je} in p5

Figure 104: The m_{je} invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.
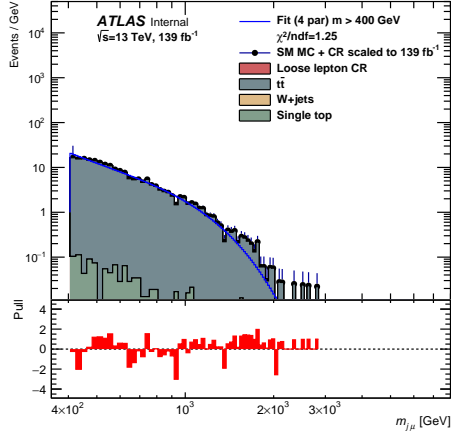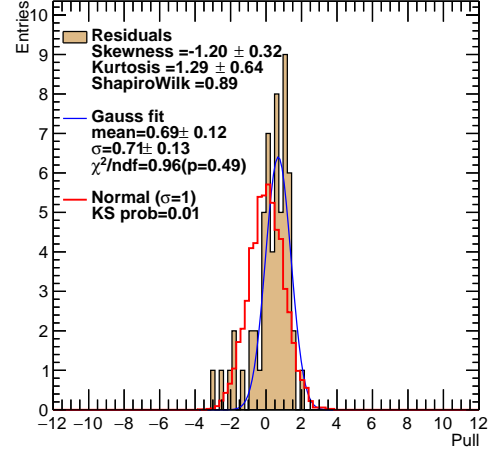
(a) m$_{j\mu}$ using MC+LE-CR, p4
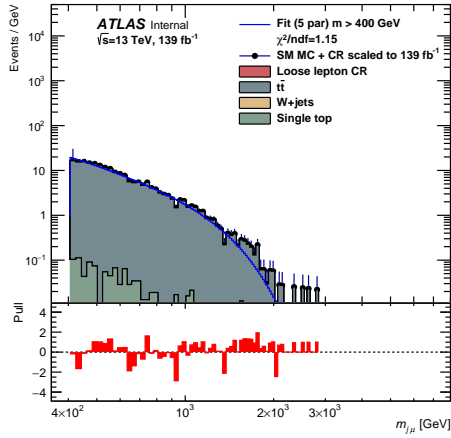
(b) pulls of m$_{j\mu}$ in p4

(c) m$_{j\mu}$ using MC+LE-CR, p5


(d) pulls of m$_{j\mu}$ in p5

Figure 105: The m$_{j\mu}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) m$_{j\gamma}$ using MC+LE-CR, p4

(b) pulls of m$_{j\gamma}$ in p4

(c) m$_{j\gamma}$ using MC+LE-CR, p5

(d) pulls of m$_{j\gamma}$ in p5

Figure 106: The m$_{j\gamma}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) m$_{be}$ using MC+LE-CR, p4

(b) pulls of m$_{be}$ in p4

(c) m$_{be}$ using MC+LE-CR, p5

(d) pulls of m$_{be}$ in p5

Figure 107: The m$_{be}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) $m_{b\mu}$ using MC+LE-CR, p4

(b) pulls of $m_{b\mu}$ in p4

(c) $m_{b\mu}$ using MC+LE-CR, p5

(d) pulls of $m_{b\mu}$ in p5

Figure 108: The $m_{b\mu}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

(a) $m_{b\gamma}$ using MC+LE-CR, p4

(b) pulls of $m_{b\gamma}$ in p4

(c) $m_{j\gamma}$ using MC+LE-CR, p5
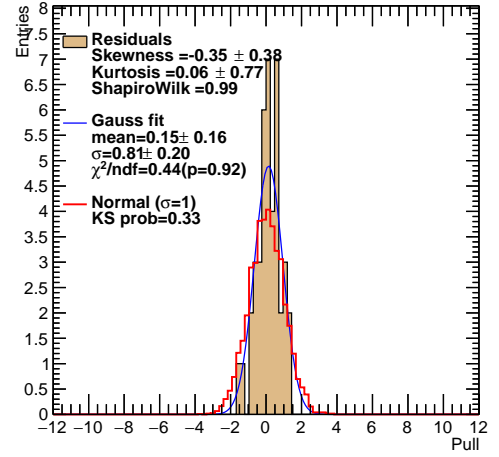
(d) pulls of $m_{b\gamma}$ in p5

Figure 109: The $m_{b\gamma}$ invariant masses with the p4 and p5 fit functions in the BSM region after the 0.1 pb AR cut is applied. The MC processes are scaled to their cross sections, while the LE-CR is used to fill the missing event rate. Pulls shown on the right.

APPENDIX J: Fit Studies on 10% of Data

The 5p fit function was shown to be the best choice from the three functions studied. However, there were few cases that strained the function and required a further step to ensure that this 5p fir function is indeed the proper choice. It was recommended to select randomly a 10% data sample and scale it by 10 to reproduce a realisitc event rate while also smoothing the distribution. The Savitzky-Golay filter [72] was used for smoothing. Figures 110-112 show the smoothed and extrapolated 10% of real data. Good agreements with the 5p fit function is observed, confirming to biases in the masses after the 10 pb AR.
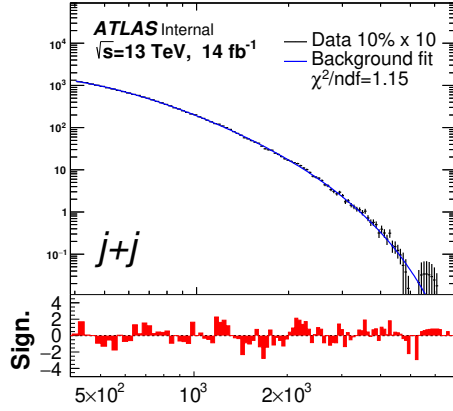
(a) jj 10% data, 5p fit
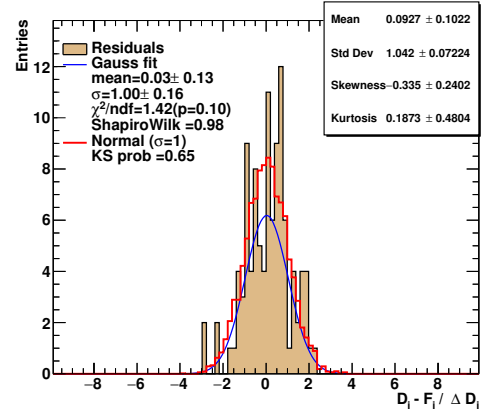
(b) pulls of jj in 5p fit

(c) jb 10% of data, 5p fit

(d) pulls of jb in 5p fit

(e) bb 10% data, 5p fit

(f) pulls of bb in 5p fit

Figure 110: Extrapolated invariant masses in 10% data using the 5p fit functions in the 10 pb AR cut. The data was smoothed using the Savitzy-Golay filter. Pulls shown to the right.
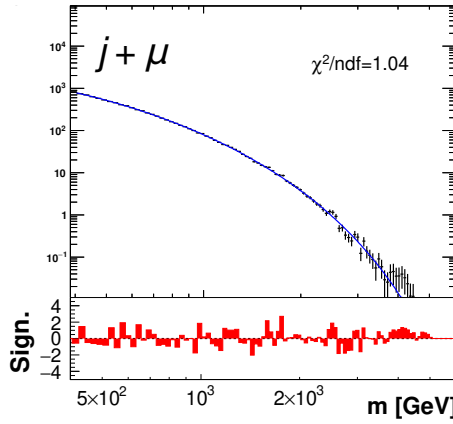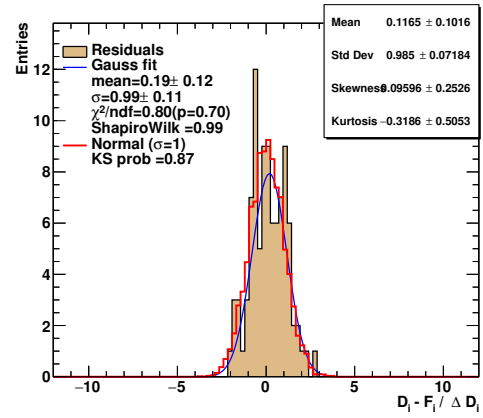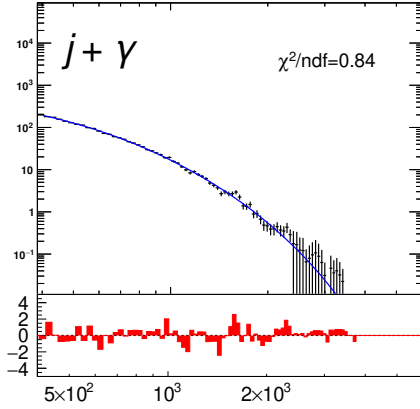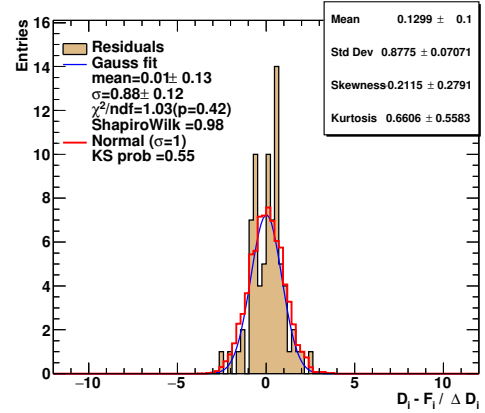
(a) je 10% data, 5p fit

(b) pulls of je in 5p fit
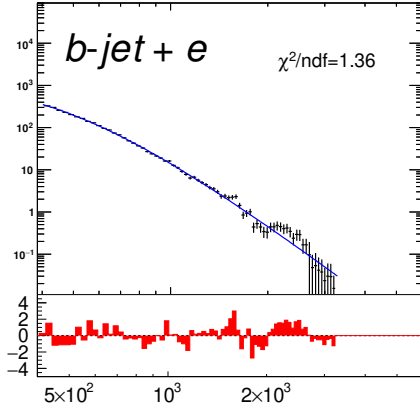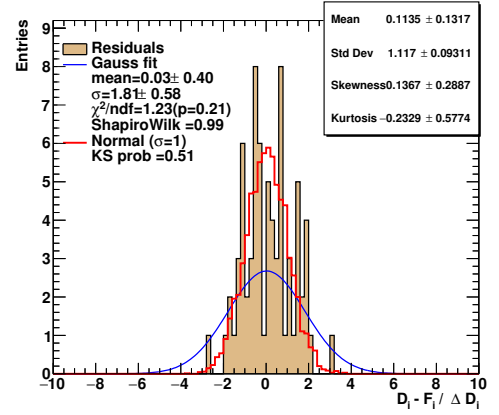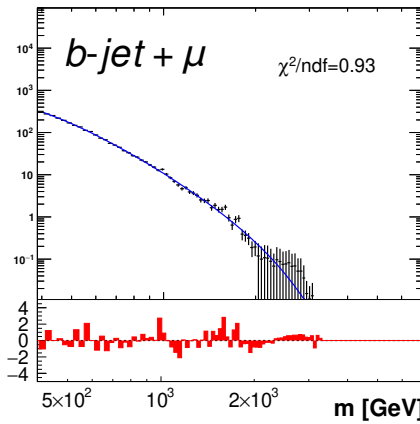

(c) jm 10% of data, 5p fit

(d) pulls of jm in 5p fit

(e) jg 10% data, 5p fit

(f) pulls of jg in 5p fit

Figure 111: Extrapolated invariant masses in 10% data using the 5p fit functions in the 10 pb AR cut. The data was smoothed using the Savitzy-Golay filter. Pulls shown to the right.
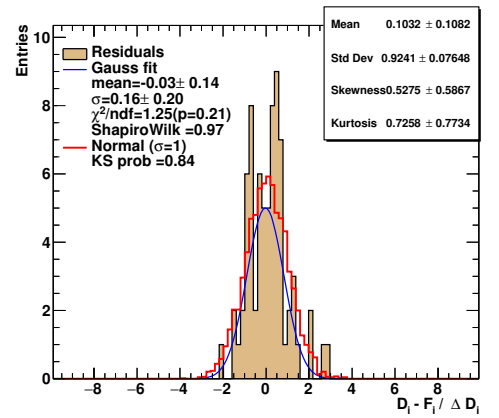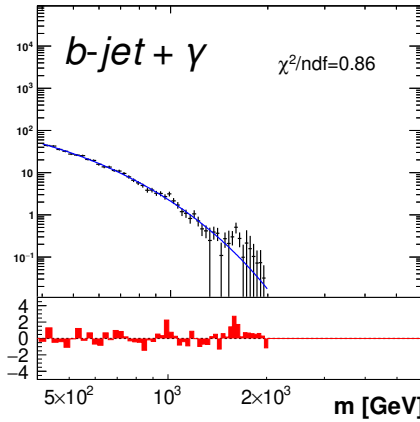
(a) bje 10% data, 5p fit
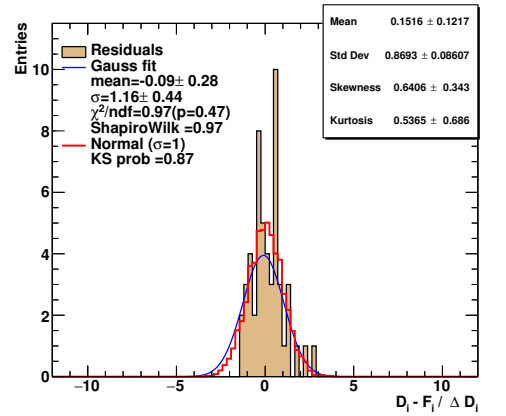
(b) pulls of be in 5p fit

(c) bm 10% of data, 5p fit

(d) pulls of bm in 5p fit

(e) bg 10% data, 5p fit

(f) pulls of bg in 5p fit

Figure 112: Extrapolated invariant masses in 10% data using the 5p fit functions in the 10 pb AR cut. The data was smoothed using the Savitzy-Golay filter. Pulls shown to the right.

VITA

Jacob E. Crosby

Candidate for the Degree of

Doctor of Philosophy

Dissertation: SEARCHES FOR NEW PHYSICS USING UNSUPERVISED MACHINE LEARN-
ING FOR ANOMALY DETECTION AT THE ATLAS DETECTOR AND THE
DEVELOPMENT OF PARTICLE IDENTIFICATION ALGORITHMS FOR THE
HL-LHC

Major Field: Physics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Physics at Oklahoma State Uni-
versity, Stillwater, Oklahoma in May, 2024.

Completed the requirements for the Master of Science in Physics at Oklahoma State Uni-
versity, Stillwater, Oklahoma in 2023.

Completed the requirements for the Bachelor of Science in Physics at Michigan State Univer-
sity, Lansing, Michigan in 2019.

Professional Memberships:

American Physical Society, European Council for Nuclear Research