University of the Witwatersrand
Faculty of Science

School of Physics

Masters Thesis

submitted for the degree of

Master of Science

What Are *B*-jets and What Can We Do With Them?

A beauty-ful exploration of *b*-jets

by

Wandile Nzuza, BSc (Hons) in Physics
Student Number: 2096716

Submission Date: April 17, 2025
Supervisor: Prof Deepak Kar

Acknowledgement

April 17, 2025

I dedicate this work to all the Black women that came before me and made this opportunity possible. I promise to continue paving the way forward. A special mention to my ancestors, who continually watch over me and put me in rooms I could have only imagined to be in. I hope they are proud. Thank you to my parents, Gugulethu Mtshali and Vusi Nzuza, for the support. My sister, Zinhle Nzuza, who always had words of encouragement for me! Continuously reminding me of how deserving and hard-working I am. To my best friends, Homestead Ekhaya, I owe you guys the world for the unconditional love and support I get. To the late-night phone calls, check-ins and messages that got me through, thank you! Thank you to my supervisor and our research group for the love and support I have been shown from the beginning. It really means the world to me! All I can say is it really does take a community. Lastly, I am proud of myself. The hard work I continually put in always pays off at the end. To the 12-year-old me who would daydream of being in a better place and doing what I enjoy most, I just want to say we have made it! You are safe, loved and appreciated. All the things you always wondered if you could do, you have done them and exceeded beyond that.

Contents

Abstract

In the Standard Model of particle physics, the bottom quark is one of the six quarks. It belongs to the third generation of quarks, alongside the top quark. The b-quark has a charge of -$\frac{1}{3}$ of the elementary charge, making it negatively charged, and a mass of 4.2 GeV significantly greater than the up and down quarks, but smaller than that of the top quark. At the LHC, two protons are accelerated close to the speed of light in a head-on collision and as a result there are final states with thousands of hadrons. This means that there are jets everywhere. Jets are not fundamental objects observed in the final state but are a cone-like structure constructed to help deal with the large number of hadrons in final states. Jets containing b-hadrons are known as b-jets. The reason that final states with b-jets are of interest is because they result in much less ambiguous collider signatures which can assist in reducing the background significantly. The results presented are from various studies done using b-jets. We started off with a study of bottom-quark-philic semi-visible jets, followed by a measurement of b-jet cross-section as a function of missing transverse momentum using data collected by the ATLAS detector in Run 2 of the LHC and lastly two Performance Studies done of heavy-flavour tagging in ALICE and ATLAS experiment at the LHC.

Chapter 1

Introduction

In this Chapter, a discussion on the Standard Model is presented. All the different components that make up the Standard Model-the quarks, leptons, gauge bosons, and Higgs boson-are explained, and how all components work together to make the Standard Model such a success. The shortcomings of this theory are also explored. Concepts such as dark matter and CP violation, which can not be explained using the Standard Model, are discussed. Lastly, a discussion on how the Standard Model is simulated using various Monte Carlo event simulations in Chapter 1.3.

1.1 The Standard Model

The Standard Model (SM) [1] describes interaction between fundamental particles and nature. It is an observation-based theoretical framework that makes assumptions and fits new data [2]. The SM consists of fermions and bosons. All matter that exists in the universe is made up of fermions [2] according to the SM. The fermions have very specific characteristics and interact through fields. The bosons are particles that are in charge of facilitating interactions with these fields [2]. Nature has four different kinds of interaction fields [1]: electromagnetic (EM), gravitational, strong, and weak. These four fields mentioned are response for various properties and behaviour of matter. They vary in strength and range.

Notably, the gravitational field has no effect on the scales that particle physics probes [2]. Because of this, the SM uses quantum-gauge field theories to explain strong, weak, and EM forces [1]. It can be simpler to consider bosons as mediators of the interactions since they are linked to the fields through which fermions interact, as was previously mentioned. This means that when these particles come into contact with one another, bosons are interchanged. For example, in the EM field, fermions with an electrical charge interchange a massless photon as they interaction [1]. The interaction between the weak field and fermions is mediated by neutral Z and $W^{\pm}$ bosons. Like photons, gluons are massless particles that act as mediators in the interaction between fermions and the strong field. This property influences the range of the interaction, which is why it is mentioned that gluons and photons have zero mass, the neutral Z boson, and the $W^{\pm}$ have non-zero mass. Strong and electromagnetic interactions should have an infinite range, whereas weak interactions have a short range. The way that EM and strong force approach infinity is where they differ. Another crucial feature of the SM is the Higgs [3, 4] mechanism, which explains how particles acquire mass. The Higgs field permeates all of space, and particles in the SM interact with it in varying strengths. The photon remains massless because it does not interact with the Higgs field. Without the Higgs field, the W and Z bosons would be massless, and the weak interaction would have an infinite range, just like EM.

As stated earlier, the SM was built on the foundation of the unification of these interactions. It is noteworthy that these quantum field theories explain these interactions, which were previously described as a theory of how fields come into contact with one another [5]. Some of the theories are only relevant when the individual components of the SM interact with one another. SM must reconcile and be consistent with special relativity and quantum mechanics, which result in the development of quantum field theory, if it is to be formulated in accordance with relativistic quantum theories. The elements in the SM are described by quantum field theories such as quantum electrodynamics (QED) and quantum chromodynamics (QCD). First, QED, which combines electromagnetism and quantum mechanics, describes electromagnetic interactions. According to its description, the photon mediates all interactions between charged particles. This indicates that photons are exchanged when charged particles come into contact with one another. QED describes the weak and electromagnetic interactions that govern leptons. Similarly, particles with colour charge can be described by QCD since they are subject to strong interactions. Electric charge is comparable to the idea of colour charge. Similar to how electric charge is conserved in all interactions in QED, colour charge is also conserved in QCD. The principle of symmetry must be mentioned in order to fully comprehend how the SM was framed [2]. Symmetry and symmetry breaking, which will be briefly discussed later, are at the heart of the notion that SM may be incomplete.

This leads us to the individual particles that comprise the SM. As shown in Figure 1.1, the SM can be summed up in a table that contains fermions and bosons. The format of the table is similar to the periodic table of elements [1]. With an intrinsic spin value of $\frac{1}{2}$, fermions are particles that follow Fermi-Dirac statistics and the Pauli Exclusion Principle [1]. Quarks and leptons are the two kinds of fermions that are present in the SM. The characteristic that distinguishes leptons from quarks is that, if they are charged, they interact through weak and electromagnetic interactions [2]. However, quarks can interact through weak, electromagnetic, and strong interactions. There are exactly three generations of fermions in the SM, it is still unclear why there are specifically three and not more or less.

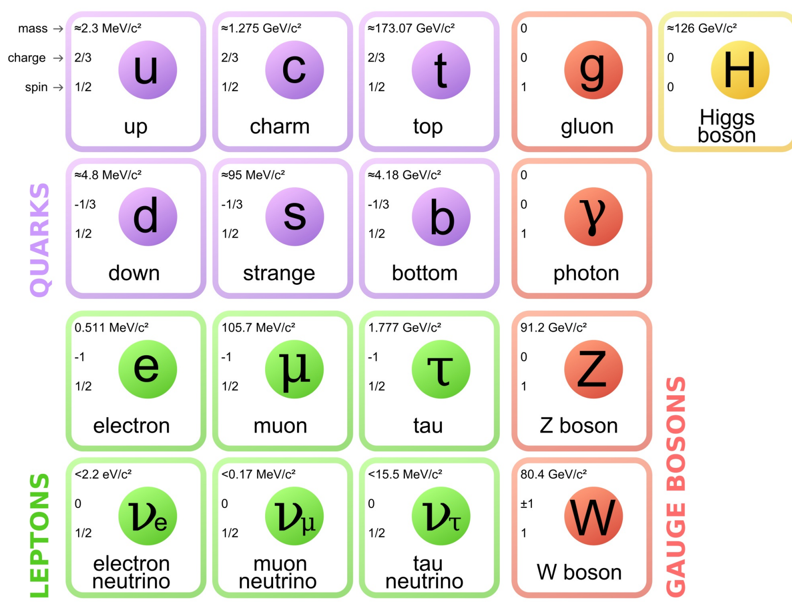

Figure 1.1: A table form of the Standard Model sourced from [6]. The scalar bosons are displayed in yellow, the vector bosons in red, the leptons in green, and the quarks in purple.

Electron, muon, tau (τ), electron neutrino, muon neutrino, and τ neutrino are the six flavours of leptons. The electron, muon, and τ are the three of the six leptons with charges of ± 1. The electron is the lightest of the three charged leptons in the SM, while τ is the heaviest. They are found in increasing order of mass. In contrast to the electron, the muon and τ are unstable, meaning they only last a short time before decaying into lighter particles [5]. Three complementary neutral neutrinos accompany the charged leptons [2]. Neutrinos are nearly massless particles that interact weakly. They have been shown in experiments to pass undetected through a variety of calorimeters and detectors. When the existence of neutrinos was first hypothesized in 1930 [1], β decay had missing energy and angular momentum. Similarly, there are six flavours of quarks: charm (c), up (u), down (d), bottom (b), strange (s), and top (t). Complementary anti-quarks accompany the quark flavours as well. The fermions are separated into three generations, as previously mentioned. The first generation consists of the quarks u and d, the second generation of quarks c and s, and the third generation of quarks t

and b. This also holds true for leptons. It is important to remember that third-generation fermions are heavier and typically created in encounters with extremely high energy. Since they are the lightest quarks, the u and d quarks are the fundamental particles found in ordinary matter.

The strong force together with QCD, asserts that no particle with colour charge can exist in isolation. Together they are responsible for the confinement of quarks in nature [7]. Red, blue, and green are the three colour charges of quarks. Additionally, these three colours have anti-colours [2]. Quarks must combine with one another in a compound system that can span more than one fm, which is why they cannot be seen as free particles [2]. This is called colour confinement. Colour confinement involves colour charges which are a representation of quantum numbers [7]. The quarks that form compound arrangements are referred to as hadrons. Baryons and mesons are the most basic compound systems where quarks can be seen. With a final quark number of three [2], baryons in particular are fermions, which are particles with half-integer spin values. However, mesons have a final quark number of zero [2], which means that they are bosons because they have integer spin values. Ultimately, hadrons must become white or colourless in order to survive in the natural world.

In later Chapters, there is mention of heavy flavour (HF) tagging which is implemented to identify the type of quark (b- or c quark) within hadrons by exploiting the decay properties of b- and c-hadrons. As mentioned, the bottom quark is part of the third generation fermions making it a few magnitudes heavier than quarks found in the SM, earning it's the name, heavy-flavour quarks. The charm quark is part of the second generation of fermions but is also considered a heavy-flavour quark, whilst the rest of the quarks, excluding the top quark, are grouped as light quarks. The reason that the top is excluded in HF tagging, even though it is the heaviest quark, is because it decays before it undergoes hadronisation. B-hadrons are hadrons that contain a bottom quark, have a lifetime of 1.5×10^{-12} s and the c-hadrons, hadrons that contain charm quarks, have a lifetime of 1.1×10^{-12} s. The point where the hard scattering of the protons takes place is known as the primary vertex as shown in Figure 1.2. The lifetime means the heavy-flavour hadrons travel a certain distance before they decay into other neutral and charged hadrons resulting in a secondary vertex as seen in Figure 1.2.

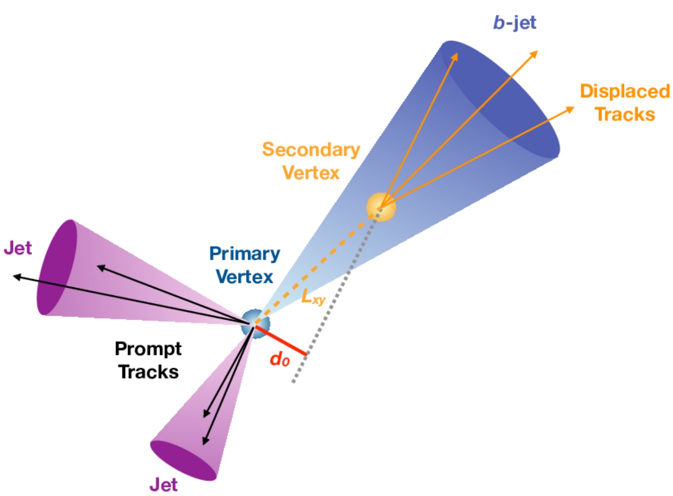

Figure 1.2: A diagram showing the structure and properties of a b-jet.

Looking at the vector bosons, the photon, gluon, W, and Z bosons, have no colour charge and an intrinsic angular momentum of 1. The non-zero mass neutral Z and $W^{\pm}$ bosons were discovered in 1983 [2] at CERN. The Higgs boson is the only scalar boson that has zero intrinsic angular momentum and no charge. As mentioned, bosons are an essential part of fundamental interactions in the SM [5] and obey Bose-Einstein statistics. Similarly to quarks, gluons have colour charge, which means they also have anti-colour charge. This means that gluons cannot be seen in isolation due to the colour charge. Instead of forming uncoloured states, like how quarks would, they can form eight possible combinations.

In the bosons mentioned already, none of them have addressed how particles obtain their mass. The Higgs field was theoretical proposed to address the absence of a massless, charged field [5] that would in theory, facilitate interactions, similarly to the other bosons. In contrast to other bosons, the Higgs field was theoretically proposed to address the lack of a massless, charged field [5] but similar to other bosons, it would facilitate interactions. A Toloridal LHC Apparatus (ATLAS) experiments [4] made the discovery of the Higgs boson in 2012, shortly after, Compact Muon Solenoid (CMS) [3] followed suite. When leptons and quarks interact with the Higgs field, they gain mass due to connecting spontaneous symmetry breaking with the principle of gauge invariance [5]. The reason behind why certain fermions are heavier, is directly linked with how strongly certain fermions interact with the Higgs field. It still remains unclear as to why this holds true.

1.2 Shortcomings of the Standard Model

There are numerous unanswered questions when learning about the various elements that make up the SM. Which convinces us that the SM is only valid to a certain extent. Despite the huge success of the SM for being in agreement with data, it is reasonable to assume that it might be incomplete. In Chapter 1.1, it was mentioned that the three fundamental interactions are unified by the SM, with gravity remaining a mystery. This leads us to the main question: why is gravity so inconsequential and unexplainable by existing field theories?

When studying the SM, concepts such as neutrino mixing and mass are not as thoroughly understood when compared to other postulations that can be explained by the SM. Based on what is currently understood about neutrino mixing and mass, the number of electron neutrinos that reach Earth is substantially lower. There should be significantly more electron neutrinos observed on Earth when accounting for the reactions that take place in the Sun [1]. This indicates that there exists a process that takes place between the Sun and Earth [1] that results in fewer electron neutrinos reaching Earth. Neutrino mixing is thought to be the process. This process explains the ability of neutrinos to change between flavours. As briefly mentioned, the Higgs mechanism provided insight on particle masses. The tiny masses of left-handed neutrinos are attempted to be explained by the Higgs mechanism. If the unification of all forces can be reached [1], there is a possibility for this neutrino mixing and masses to be better understood. This also means right-handed neutrinos with significantly higher mass could be introduced.

CP violation is the term to describe the observation that quarks in the SM do not remain the same under charge and parity operations [2]. It is still unclear how this invariance is violated. It was first identified in 1964, in the decomposition neutral K mesons [2], which was also observed in neutral B mesons. With the current constraints on CP violation, the SM is used to explain the existence of matter in the universe, but the absence of anti-matter in the universe today can not be explained. Assuming that matter and anti-matter were created in equal amounts during the universe's creation. The disparity between matter and anti-matter can not be explained by what is currently known about CP violation [7].

When calculating the ratio between mass and luminosity as a means to account for and reconcile the mass of the universe, there is evidently an existence of non-luminous matter called dark matter (DM). Approximately 25% of the mass in the universe is comprised of DM, which assembles to form galaxies' structural outlines and hardly interacts with known matter. To understand the nature of DM, some searches focus on the gravitational signatures that emerge from DM [8].

1.3 Simulating Standard Model Processes

Being able to simulate the SM is important because it allows for precise predictions and accurate conclusions to be made about the SM. The computational simulations model the entire complexity of SM processes from the hard scatter p-p collision to objects such as photons and electrons found in the detector. Once the SM processes are simulated, a list of particles that would be found in the final state is provided with their four momenta. These simulations utilize a range of techniques that are generally classified as part of the Monte Carlo (MC) method. These MC methods are applied in the lattice calculations, event generators, and perturbative N^{n}LO programs [9].

The MC method [10] is a versatile statistical approach employed to approximate complex integrals through random sampling. The SM processes that are simulated can also be represented using Feynman diagrams. These diagrams are a visual representation of the interactions between elementary particles. When the amplitudes coming from Feynman diagrams are integrated in all orders, it is referred to as the scattering amplitude [7]. The probability of a particular process happening, also known as the cross-section, is calculated by squaring the scattering amplitude. MC methods are instrumental in determining observables that emerge from intricate, multi-dimensional phase spaces, such as the cross-section of processes involving multiple particles.

MC event generators are software frameworks that utilize the MC methods to implement various physics models. Generators such as `PYTHIA` [11, 12], `SHERPA` [13], `HERWIG` [14, 15], `POWHEG-BOX` [16, 17, 18] and MadGraph_aMC@NLO [19] are designed to simulate the complete sequence of *p*-*p* collisions, integrating theoretical frameworks from QCD, electroweak interactions, hadronisation, multiple parton interactions (MPI), parton distribution functions (PDFs), initial- and final-state radiation (ISR/FSR),hadronisation, particle decays. These generators merge matrix element (ME) computations with parton showers (PS) and phenomenological models to create realistic event simulations. In Chapter 3, `PYTHIA` and MadGraph_aMC@NLO are used and in Chapter 4 `PYTHIA` and `SHERPA` are used.

Figure 1.3 is a visual representation of different generators that are fed into MC event generators; the events simulated are also used in detector simulations. The steps for processing data collected from the detector are discussed in Chapter 2. The simulation process of the SM is directed by QCD, as mentioned, especially the factorisation theorem. The factorisation theorem of QCD allows the separation of short-distance (perturbative) physics from long-distance (non-perturbative) effects, facilitating the development of event simulations [20].

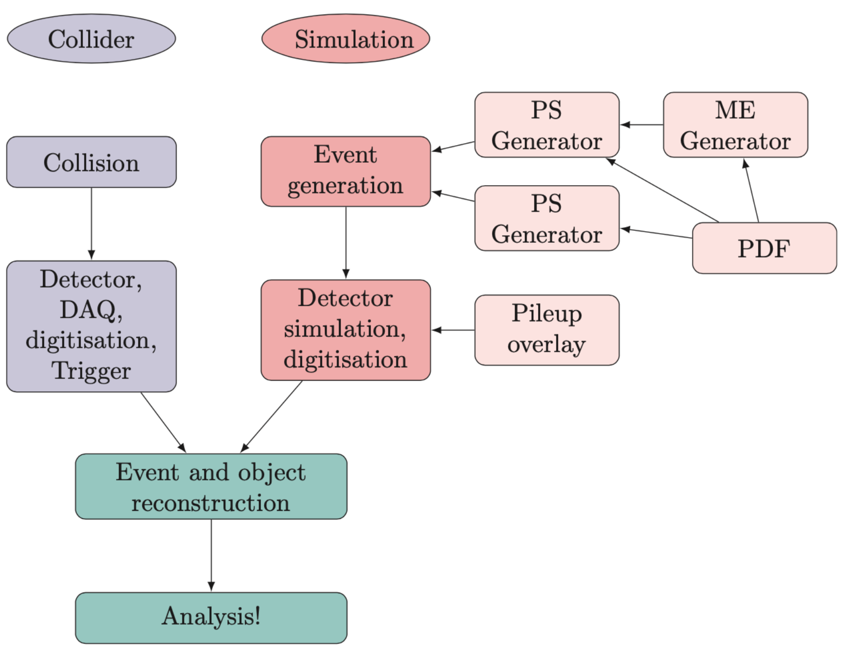

Figure 1.3: Diagram showing the steps for simulating events, compared with the processing steps for the data taken from [7]. The event and object reconstruction set-ups are identical.

One significant outcome of the theorem is the capacity to derive parton-level information from the proton through parton distribution functions (PDFs) [21, 22], which serve as the basis for any MC event generator, as shown in the Figure 1.3. During the hard process, the fraction of energy that the incoming partons have is known as PDFs. It accounts for the probability of a parton having a certain fraction of energy from the proton. It is presented as $f_i(x, \mu_F^2)$, which is the probability that parton i is carrying a fraction x of the proton's energy at a given factorisation scale μ_F [22]. The factorisation scale is introduced to distinguish non-perturbative dynamics from the perturbative hard cross-section [7]. PDF sets that are used for analysis, which can be at LO, NLO, and beyond, are produced by a few different theoretical-experimental groups. The PDF used in Chapter 3 and 4 is Neural Net PDF (NNPDF) [23]. It is available on the interface Les Houches Accord PDF (LHAPDF) [24, 25].

Following PDFs providing information on the energy distribution in incoming partons, the subsequent steps involve calculating the likelihood of partons engaging in interactions as described by SM processes; this is done by matrix element (ME) generators. The role of ME generators is to handle interactions between partons using perturbative QCD [19]. ME generators model the hard scattering. An example of ME generators is `POWHEG-BOX` [16, 17, 18] and MadGraph_aMC@NLO [19]. The generator MadGraph_aMC@NLO is integrated with AMC@NLO, which calculates ME at NLO. The ME generator simulates processes such as $q\bar{q} \to Z$ with high precision, frequently incorporating higher-order corrections.

The precision of an ME calculation is often described in terms of its perturbative order, as mentioned, it can incorporate higher-order corrections. At leading order (LO), the process is represented by the simplest Feynman diagrams, typically with the fewest vertices and no loops [26]. Going to next-to-leading order (NLO) adds one-loop corrections, while next-to-next-to-leading order (NNLO) includes even more refined corrections, such as two-loop diagrams [26, 27]. By adding one or more loops to the Feynman diagram, it increases the precision in the cross-section of a particular event.

Independent of the perturbative order is the final-state multiplicity, which is frequently characterized by multi-leg ME. A process is referred to as a multi-leg process when it is, for example and not limited to, a $2 \to 3$ or $2 \to 3$ process. While a $2 \to 2$ process involves p-p collision with two final state partons. It is important to note that a multi-leg process can be at LO; it just includes more final state particles. In contrast, a higher-order process (NLO, NNLO) involves more precise corrections to a given SM process, not necessarily more legs. The Figure 1.4 shows a simple Feynman diagram to depict the difference between a multi-loop and multi-leg process.

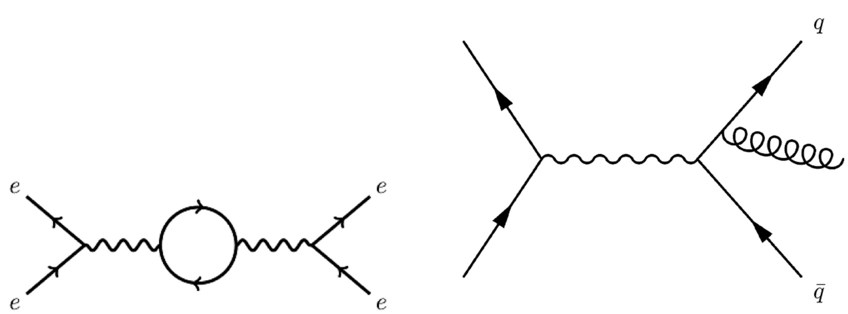

Figure 1.4: Illustration of Feynman diagram. The Feynman diagram on the left-hand side depicts a multi-loop process whilst the diagram on the right-hand side depicts a multi-leg process. The Feynman diagrams are taken from [7]

However, ME calculations are not sufficient to describe all aspects of the event—particularly when it comes to the evolution of p-p collision, especially soft or collinear emissions [28]. This gap is filled by parton shower (PS) [29, 30] generators. PS simulate the radiation of soft and collinear gluons and photons before and after the hard scattering [28]. These emissions are not captured by ME calculations, as mentioned, but are crucial for reproducing realistic jet structure and multiplicities [7].

The PS process estimates the direction and energy of the hard scatter and adds branches to the hard scatter to create the full event structure. With each parton branching, there is a decrease in energy scale, and the new branching happens at energy scales lower than the previous [7]. The PS process continues until the energy scale of the partons reaches colour neutral hadrons. With each branching, there is a decrease in energy scale. The parton branches can develop before or after the QCD interaction, which are referred to as initial-state and final-state showers [7].

In Figure 1.3, it is shown that PS generators can be interfaced with ME generators. The reason for this is ME generators are used for simulating the hard scattering, and they describe individual hard objects better. On the other hand, PS generators describe the evolution of the event structure. Since PS and ME generators both simulate radiation in overlapping kinematic regions, a matching and merging procedure must be implemented to avoid double counting [31]. Both ME and PS generators can not run simultaneously, as it will result in double counting of parton emission [7]. A procedure called matching and merging is implemented to ensure that the parton emissions are not double counted. The matching and merging scales are defined using either p_{T} or energy of a jet to avoid overlap [7]. In Chapter 3, the MLM jet matching scheme is used [32]. This matching scheme ensures that jets from PS are matched with jets from ME partons. The event is kept if the PS jet matches a patron in the ME event.

Once the perturbative part of the event is modelled, the next step is to simulate the non-perturbative transition from partons to colour neutral hadrons [33]. Hadronisation/fragmentation takes place after showering; it is responsible for converting partons to hadrons. There are two phenomenological models used, known as the cluster model [34, 35] and the Lund string model [36, 37]. These hadronisation models describe how quarks and gluons, which cannot exist as free particles, transform into bound hadronic states. The event generator `PYTHIA` makes use of the Lund string model, and `HERWIG` uses the cluster model. Depending on the lifetime of the hadrons formed, they may further decay. Even though strong interactions dominate at hadronic colliders, electroweak interactions often have non-negligible contributions [7]. When doing measurements and searches, estimating electroweak corrections becomes important.

Despite the sophistication of MC event generators, such as `PYTHIA`, `SHERPA`, and `HERWIG`, as mentioned, they are based on physics models that have free parameters that control different parts of the model. These parameters can be set independently from each other, but it can be complicated since they can end up being correlated. These parameters model processes such as parton showers, MPI, hadronisation, and the underlying event. To ensure these components realistically reflect what is seen in data, a process known as tuning is applied [38]. Tuning is a process implemented to set these parameters simultaneously as a group using data distributions that are sensitive to specific physics aspects. For both Chapter 3 and 4, A14 [39] and A3 [40] tune are used. With tuned parameters in place, the generator output provides a more accurate description of the underlying physics. These events are then passed into detector simulations, which model how the particles interact with various materials and components of the detector.

When it comes to detector simulations, simulations such as `GEANT4` [41] are used. Once the events are generated using MC event generators, they can be passed through detector simulators, which mimic the different layers in the detector using MC principles. These simulations model how different materials in the various layers of the detector would interact with specific particles. The various layers are made of material that has magnetic and electric fields present; as the particles move through each layer, they can deposit energy on the material, which can create a shower of new particles that move independently through the detector. Electromagnetic and hadronic have to be modelled differently due to processes such as bremsstrahlung and multiple Coulomb scattering [7]. These interactions have to be simulated until the particle reaches the end of the detector or below a certain energy threshold [7].

Once the events have been generated, it is important that despite different MC event generators being used, they must have a consistent output format. PS generators save the events in HEPMC [42] format. Generators producing parton-level output save events in Les Houches Event File (LHEF) [43] format. Different experiments use their own internal format based on these. Once they are digitised, they can be used in analysis.

Chapter 2

Experimental Set-ups: ATLAS and ALICE

In this Chapter, both the ATLAS and ALICE detectors are discussed. Starting from the coordinates used to describe various particles to the infrastructure and the different materials in the distinctive layers. Followed by a description of the trigger systems in ATLAS and ALICE, which are designed to filter and select potentially interesting events from the large volume of data generated at high-energy collisions. Both experiments use a multi-level trigger system, with hardware-based triggers at the first level to quickly reduce event rates, followed by software-based triggers at higher levels to perform more detailed analysis and event selection. Lastly, details on the object reconstruction for the measurement and performance studies in the later Chapters are presented.

2.1 Detector Infrastructure

The Large Hadron Collider (LHC), located on the border between France and Switzerland, also referred to as the LHC ring, has a circumference of 27 km. It collides protons close to the speed of light to test the predictions of SM and to possibly uncover new particles, but it can also collide heavy ions in Pb-Pb collisions. In Figure 2.1, all of the experiments that are found in the LHC ring are shown.

The ATLAS detector [44] is a multi-purpose particle detector designed to uncover a wide range of physics not answered by the SM and also discover new particles. As mentioned, it collides two protons close to the speed of light. The two protons are obtained by applying an electric field to hydrogen gas, which separates the atoms in the gas. The protons obtained from the hydrogen gas are accelerated in stages by a series of circular and linear accelerators until they are injected into the LHC ring. Once the protons are in the LHC ring, they are injected in both clockwise and anti-clockwise directions in bunches. The detector can collide 2808 bunches of protons per beam every 25 ns. The beams are kept in trajectory by superconducting dipole magnets; these magnets squeeze the protons together so when they enter the detector as a narrow beam. After the hard scattering of the protons, the particles emitted are separated using dipole magnets.

The particles emitted as a result of the collision are probabilistic. Quantum mechanics enforces that there is no way to predict what the outcome of any specific collision will be with certainty, so this means that there is no exhaustive list of possibilities. How often a particular process occurs is known as cross-section. The cross-section of each process can be measured using information from the colliding beams. Some other quantities that can be calculated using the colliding beam are probability of collision, total number of collisions, collision rate, and the actual number for a process. The probability of a collision is calculated as

$$P_{collision} = \frac{\sigma}{A_1/n_1} \tag{2.1}$$

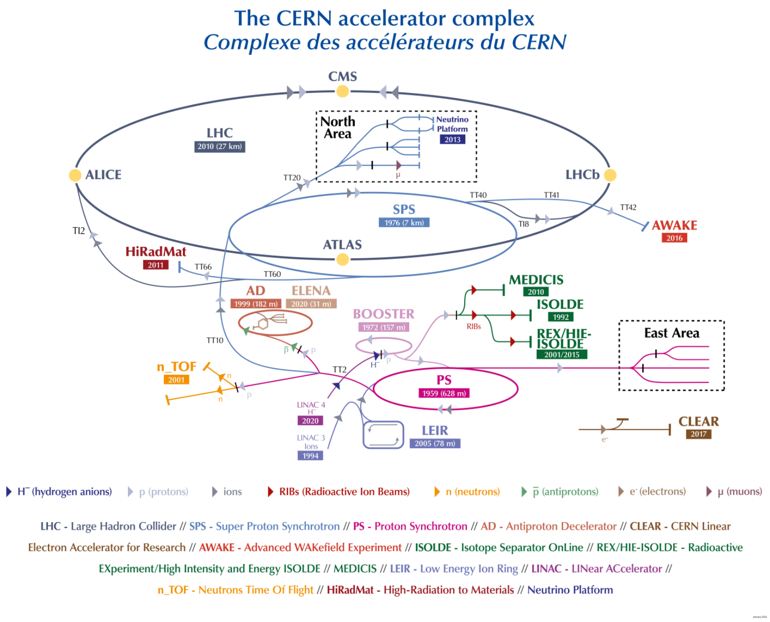

Figure 2.1: A diagram showing the LHC ring located between France and Switzerland. The various experiments that are CERN are labelled along the ring

The σ is the effective area of one particle, A_1 is the total area of the beam 1 and n_1 is the number of particles a beam 1. The total number of collision is given by computing:

$$N_{collision} = \frac{n_2\sigma}{A_1/n_1} \tag{2.2}$$

The collision rate is given by:

$$R_{collision} = \sigma f \frac{n_1 n_2}{A_1} \tag{2.3}$$

The f in the equation is known as frequency of collision, and $f\frac{n_1 n_2}{A_1}$ is also known as instantaneous luminosity. Luminosity measures how many particles are put in position to collide. Measuring the luminosity is important because it is closely related to the cross-section, and it has the units $\frac{1}{area}$. The smaller the value of A_1, the higher the collision rate will be because it means the particles are tightly packed together. The actual number for a process is given as:

$$N_{process} = \sigma_{process} \int L dt \tag{2.4}$$

The number for a process depends on both the physics and the accelerator. The physics accounts for the physical process being considered, and the accelerator accounts for the design of the detector. The $\sigma_{process}$ represents the physics, and $\int L dt$ represents the accelerator. In the equation $\int L dt$, it is also known as the integrated luminosity, which is just the instantaneous luminosity integrated over a specific time period.

The LHC has a discrete interval data-taking period known as a Run. In between each Run, there are long shutdowns where the detector is maintained and upgraded. During the shutdowns there are no running beams, but the detector is still switched on for calibration purposes. Run 1 was from 2009 to 2013; the first long shutdown was in 2014. Run 2 was from 2015 to 2018, and the second long shutdown was in 2019, and Run 3 started in 2020 and is still ongoing currently. At the end of Run 3, the centre of mass will be 13.6 TeV. During the shutdowns, equipment such as magnets is upgraded, which means that the luminosity also increases with each run. As the luminosity increases, there is also an increase in pile-up. Pile-up is the additional proton-proton collision that occurs within the beam bunches that contaminate the event with extra particles. The detector resolution has its own limitations; it is unable to distinguish between the interactions between the beams and single beam crossings.

When it comes to the geometry of the detector. It has a cylindrical symmetry with a near 4π coverage in solid angle aimed to cover the entire collision point [7]. So when it comes to the coordinates, only the x-y plane is relevant, and it contains both the energy and position information. Since the particles are collided close to the speed of light, relativistic kinematics are considered. The coordinates of the particle are given by a four-momentum vector (E, p_X, p_Y, P_Z), and in the centre-of-mass frame, the colliding particles have a transverse momentum of zero. It is worth knowing that the invariant mass is frame-independent and is the same as the rest mass for a single particle: $m_{inv} = E_{cm} = \sqrt{s}$. With each beam having 6.5 TeV worth of energy, the centre-of-mass energy is $\sqrt{s} = 13$ TeV. In Figure 2.2, illustrated is a sketch of the cylindrical geometry of the detector. The Cartesian coordinates are also labelled in relation to ϕ and η.

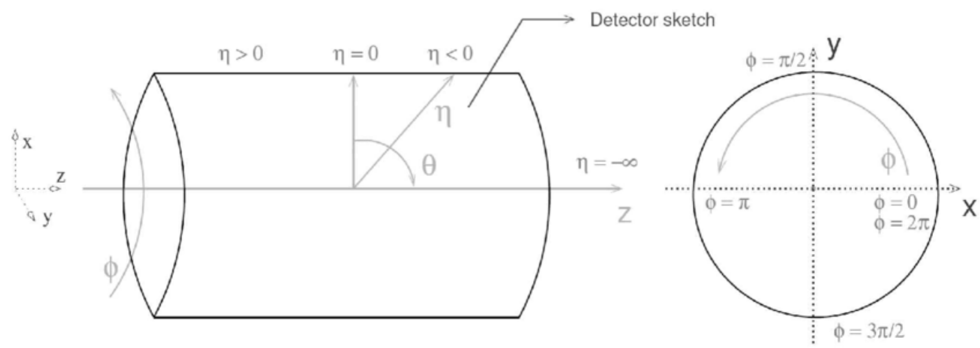

Figure 2.2: A diagram illustrating the cylindrical symmetry of the ATLAS detector. The relevant coordinates are labelled in the sketch in relation to Cartesian coordinates.

When it comes to translating the four momentum vector into cylindrical symmetry, p_Z can not be measured as it is along the beam axis, as shown in Figure 2.2 and Figure 2.3. The transverse momentum p_{T} is calculated as $p_{\mathrm{T}} = \sqrt{p_x^2 + p_Y^2}$. The η is calculated as $\eta =$ - ln tan $\theta/2$. Figure 2.4 shows the different θ values and the subsequent η value. The $\frac{1}{2}$ factor in the equation to calculate η comes from the angular coordinates, which are scaled in the identical direction, and an object at $\theta/2$ has the best resolution and same behaviour in physical space. The distance in detector coordinates is calculated as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. When describing the coordinates of a particle found in the detector, it is described as (p_{T}, η, ϕ). In Figure 2.3, the LHC ring in relation to the detector is illustrated. The IP stands for interaction point. This is where the hard scattering takes place and the centre of the detector. The Cartesian coordinates alongside the detector coordinates are shown.

The detector has 4 distinctive parts, namely the inner tracking detector, muon spectrometer, electromagnetic calorimeter, and hadronic calorimeter. In Figure 2.5, all 4 parts are labelled. Each part of the detector is designed and constructed with specific materials to optimize the detection of particles based on their properties. The inner tracking detector is surrounded by a thin superconducting solenoid providing a magnetic field of 2 T, can detect particles with a minimum p_{T} of 0.1 GeV, and has an impact parameter (d_0) resolution of 100-40 μm. It also has a pseudo-rapidity range $|\eta| < 2.5$ and consists of silicon pixel, silicon micro-strip, and transition radiation tracking detectors. Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity within the region $|\eta| < 3.2$. A steel/scintillator-tile hadronic calorimeter covers the central pseudo-rapidity range ($|\eta| < 1.7$) [45]. The endcap and forward regions are instrumented with LAr calorimeters for EM and hadronic energy measurements up to $|\eta| = 4.9$. The muon spectrometer surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. The muon spectrometer includes a system of precision tracking chambers up to $|\eta| = 2.7$ and fast detectors for triggering up to $|\eta| = 2.4$. The luminosity is measured mainly by the LUCID–2 [46] detector, which is located close to the beam pipe.

The ALICE (A Large Ion Collider Experiment) [47] is a heavy-ion collider also found in the LHC. The

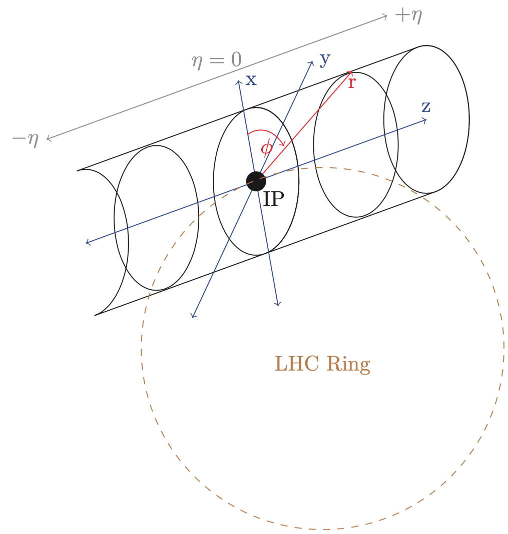

Figure 2.3: Diagram showing the LHC ring in relation to the geometry of the detector. The IP stands for the interaction point. This is where the hard scattering of the protons takes place. The x, y and z coordinates are labelled alongside the η, ϕ and r. The diagram is taken from [7].

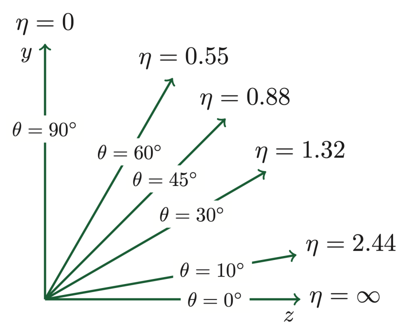

Figure 2.4: In detector coordinates there is p_{T}, η and ϕ. In the diagram illustrated the corresponding η values are given in relation to the θ. The diagram is taken from [7].

focus of ALICE is to probe the strong interaction sector in the SM. With a goal of investigating quark-gluon plasma in extreme energy and temperature values. Similarly to the ATLAS detector, it can measure hadrons, electrons, muons and photons. The ALICE detector also has a cylindrical shape, as seen in Figure 2.6. The central part known as the barrel measures the hadrons, electrons and photons, with a forward muon spectrometer dedicated to measuring the muon. In Figure 2.7 a 2D annotated view of ALICE is shown. The central part covers the angles of 45° to 135° surrounded by a large solenoid magnet reused from the L3 experiment found at the LHC. The barrel contains an Inner Tracking System (ITS) which is divided into 3 planes but overall has 6 high-resolution silicon detector layers (silicon pixel detector, silicon drift detector, and silicon strip detectors), each plane with a different η and ϕ acceptance [47]. The ITS also consists of a cylindrical Time-Projection Chamber [47] with an η acceptance of ± 0.9 and a tracking radiation detector filled with Xe/CO_2 with an η acceptance of ± 0.84 for tracking in the central part of ALICE. There are three particle identification arrays (Time-of-Flight, Ring Imaging Cherenkov and Transition Radiation detectors) and two electromagnetic calorimeters (PHOS and EMCal). Ring Imaging Cherenkov and the two EM calorimeters are the only two components of ALICE that do not cover the full azimuth. The EMCal is a Pb scintillator used to detect charged particles with an acceptance of $|\eta|$ ± 0.7 and $80° < \phi < 187°$, whilst the PHOS, made of dense scintillating crystals, has an acceptance of $|\eta|$ ± 0.12 and $220° < \phi < 320°$ and detects photons and neutral mesons. The forward muon arm covers the range 2°–9° and an acceptance of $-4 < \eta < -2.4$ consisting of a complex arrangement of absorbers, a large dipole magnet, and fourteen planes of tracking and triggering chambers.

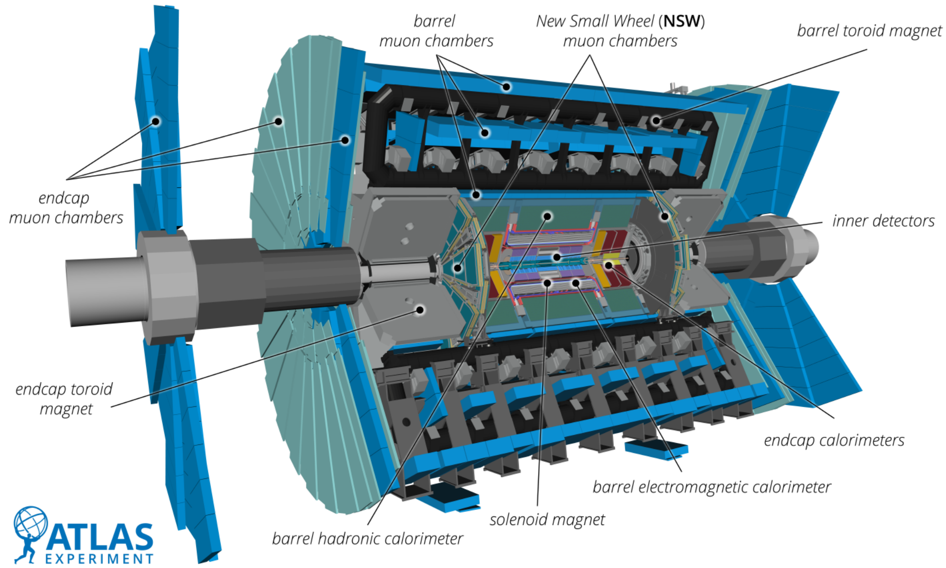

Figure 2.5: A schematic diagram of the ATLAS experiment and the distinctive layers that make up the detector and are used for particle reconstruction.

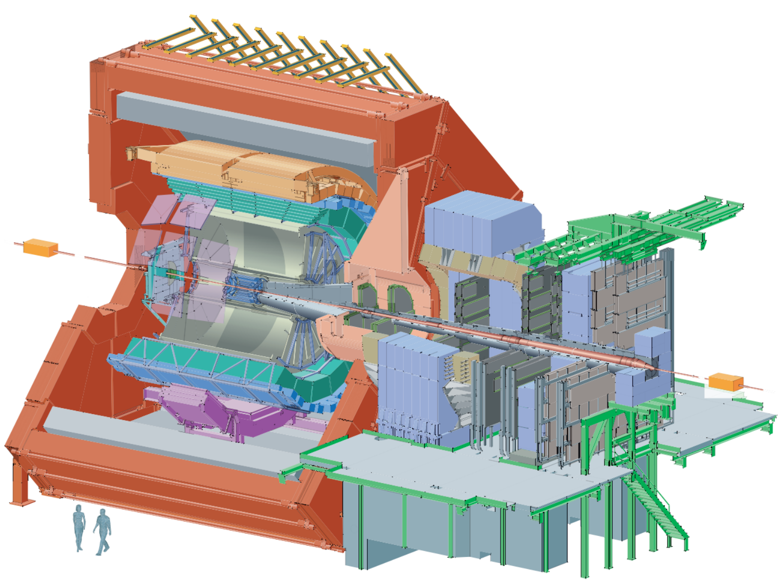

Figure 2.6: A schematic layout of the ALICE experiment. Image taken from [47]

2.2 Trigger Systems

ATLAS records roughly 300 events per second [47], with 1 fb^{-1} equating to 10^{12} proton-proton collisions at the LHC; this amounts to about 500 TB of data. The trigger is designed to select only potentially interesting events and discard the rest. It is implemented to deal with the limitation of data storage and transmission. It is implemented in stages; the output information in each level becomes input data in the next trigger level. Each trigger level has more time in order to make better decisions on which events are stored and which are discarded. Level 1 trigger has the tightest time and storage constraint and is implemented in the hardware [48]. It can only select a subset of events from the muon spectrometer and calorimeters at a rate below 100 kHz [45]. It is software-based and also reduces the event acceptance rate. The next trigger level is called high-level trigger (HLT) and deals with detailed

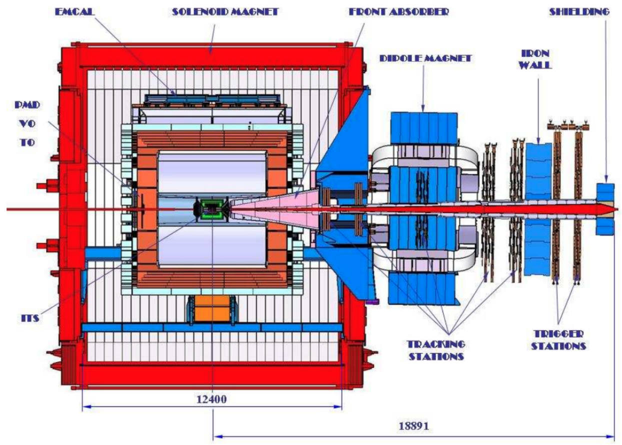

Figure 2.7: An annotated 2D illustration of the ALICE detector along the yz direction. Image taken from [47]

analysis of full event data. It can select events, for example, for 1 or 2 leptons with minimum p_{T} or jets with minimum p_{T}. A software suite, with details on the software and firmware, is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

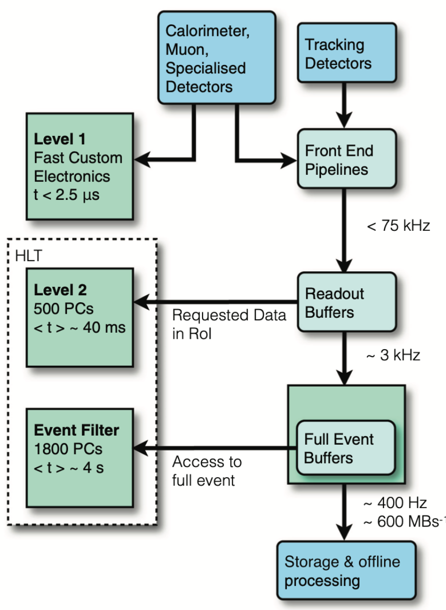

Figure 2.8: Schematic diagram of the trigger levels implemented in ATLAS experiment. Image taken from [47]

ALICE has a three-level trigger system [47]. The hardware-based triggers make use of the inner tracking system, three particle identification arrays and two EM calorimeters as inputs. The trigger also operates on different levels aligned with the different time requirements of the detectors in the experiment. There is a pre-trigger which activates the hardware-based electronics after each interaction. This operates at a timescale of less than 900 ns. Each level has time constraints and attempts to reduce the event rates depending on trigger inputs. The L0 trigger activates the tracking radiation detector, which works at a timescale of 1.2 μs; L1 at 6.5 μs; and L2 at 100 μs and is activated after the slowest detector.

The software-based trigger in ALICE is also known as HLT [47] and also works complete events to a detailed online analysis. In Figure 2.9 a visualisation is provided of the layers and structure of the HLT taken from [47]. The software-based trigger has up to 1000 multiprocessors [47] PCs which are tasked with reducing event size by either selecting to read out event information from the region of interest or compressing event information. The ALICE experiment also does heavy-ion collisions, but the runs are relatively short due to the limitation in data storage. Heavy-ion collisions require a large bandwidth of 1.25 GB/s [47] to obtain a sufficient number of events to do physics analysis. The data storage requires sophisticated optical data links and scalable architecture.

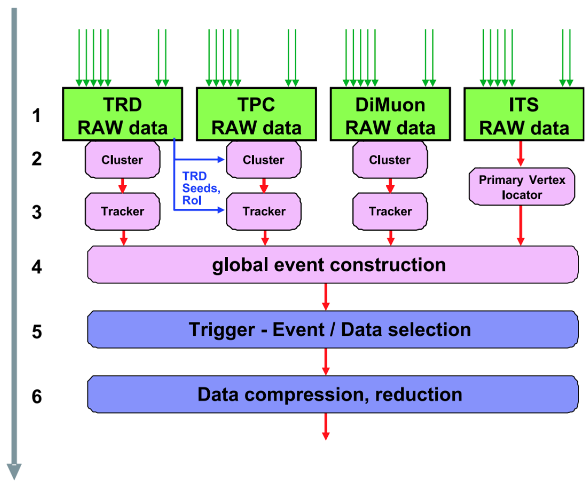

Figure 2.9: Diagram depicting the structure and layers in the software-based component of the trigger. Image taken from [47]

2.3 Object Reconstruction

The vertices are reconstructed using ID tracks. The reconstructed vertex must have at least two associated tracks with $p_{\mathrm{T}} > 500$ MeV. The primary vertex is decided based on the largest sum of the squared transverse momenta, which is computed using all associated tracks ($\sum p_{\mathrm{T}}^2$).

When two protons are accelerated close to the speed of light in a head-on collision, and as a result, there are final states with thousands of hadrons. This means that there are jets everywhere. Jets are not fundamental objects observed in the final state but are constructed to help deal with the large number of hadrons in final states. When dealing with an object such as jets, it can get complicated. A jet, as seen in Figure 2.10, is a cone-like structure that encloses hadrons. The blue jet in Figure 2.10 is known as a τ-jet signature; it contains all the final state particles that come from τ decay.

In ATLAS there are few jet clustering algorithm that can be used to construct jets. To construct fixed cone radius jets the following equations are used:

$$d_{ij} = min(p_{\mathrm{T}i}^{2\rho}, p_{\mathrm{T}j}^{2\rho})\frac{\Delta R_{ij}^2}{R^2} \tag{2.5}$$

$$d_{iB} = p_{\mathrm{T}i}^{2\rho} \tag{2.6}$$

$$R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{2.7}$$

As mentioned in a final, there will be hadrons, and each of them has four momentum vectors, which describe the p_{T}, η, and ϕ values of the particle. A jet is designed to cluster hadrons that are close to each other; equation (2.7) is used to calculate the angular distance between the hadrons found in the final state. The exponent ρ found in equations (2.5) and (2.6) controls the ordering or merging depending on its chosen value. The variable ρ can take up any value, but the values -1, 0 or 1 are typically used. The jet is formed by calculating d_{ij} and d_{Bi}. For each hadron d_{Bi} is computed, which

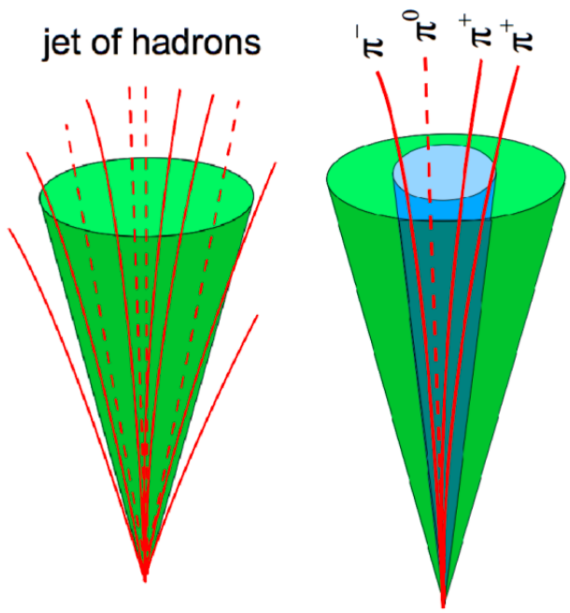

Figure 2.10: A diagram displaying how jets are constructed in hadron colliders. There is a jet which contains hadrons known as a hadronic jet and a blue jet which contains final state products of τ decay known as τ-jet

assists in figuring the minimum between $p_{\mathrm{T}i}^{2\rho}$ and $p_{\mathrm{T}j}^{2\rho}$. If the smallest value is among d_{ij} and d_{iB} for inputs i and j, and it is d_{ij}, the inputs are merged. The merging implies that the p_{T} of the hadron input i and j are added, while the position of the merged input is the p_{T} weighted average position. If the smallest value between d_{ij} and d_{iB} is the d_{iB} value, the i^{th} input is considered a jet. That particular hadron is therefore not considered an input. This procedure continues until all inputs are merged or classified into jets. Ultimately, the jet clustering algorithm works by finding the minimum distance until no particles are left.

As mentioned, the variable ρ typically takes the values of -1, 0, and 1. When $\rho = 1$, this is known as the k_t algorithm. When calculating d_{iB}, the inputs with the smallest p_{T} are clustered first in the jet, and the inputs with larger p_{T} values are added last, resulting in irregular amoeba-shaped jets. It is worth noting that the k_t algorithm represents an approximate inversion of the QCD branching process, preserving the actual clustering history. When $\rho = 0$, this is known as the Cambridge-Aachen (C/A) algorithm. For this algorithm the inputs are merged based on angular separation. When d_{iB} is computed, the exponent takes the value of zero. When ρ = -1, this is known as the anti-k_t algorithm. This is the most predominantly used jet clustering algorithm in the ATLAS experiment. Unlike the k_t algorithm, which adds the softest inputs first, in anti-k_t the inputs with the largest p_{T} values are added first in the jet. This results in the formation of roughly circular-shaped jets with an area of πR_0^2. This means that jets can be described as cone-like jets without the associated shortcomings. This is the reason behind this algorithm being favoured experimentally. The order of merging is not motivated by the QCD branching process in comparison to the k_t algorithm.

As part of the reconstruction, depending on the jet clustering algorithm, the energy deposits are gathered from the calorimeter clusters and tracks, and the information is then used to reconstruct the jets. For the measurement of the b-jet cross-section, Particle Flow (PFlow) jets were used [49]. PFlow just means a combination of tracks and calorimeter information is used to construct a jet at the particle level. The jets were constructed using anti-k_t with $R = 0.4$ [50, 51]. PFlow jets use charged and neutral constituents as inputs [49]. The charged constituents are part of the particles reconstructed from the primary vertex with p_{T} and $|\eta| < 2.5$, whilst the neutral constituents are a combination of track and calorimeter information. PFlow jets exclude muons and neutrinos in the construction of jets. The jets used in the measurement were required to be within $|\eta| < 2.0$ and have $p_{\mathrm{T}} > 30$ GeV.

For detector-level jets, the same settings and procedures were used. For the reconstruction of truth-level jet electrons and muons are not included in the jet construction. These jets are calibrated using information from calorimeter cell clusters to construct jets using the anti-k_t algorithm with $R = 0.4$.

The results from the performance study in ALICE used jets that were reconstructed using tracks from charged particles with p_{T} from 0.15 GeV to 1000 GeV, reconstructed in the central tracking detectors within the pseudorapidity range $|\eta| < 0.9$ and φ range of 0 to 2π. The jets were also assembled using the anti-k_{t} algorithm [50] with $R = 0.4$.

HF tagging is a powerful technique employed by both ALICE and ATLAS experiments to discern and isolate HF jets amidst the myriad of other particles produced during collisions. Only a few events with jets contain heavy-flavour hadrons that originate from bottom and charm quarks through hadronization/fragmentation [52]. The purpose of HF tagging is to identify the type of quark (b- or c quark) within hadrons and label the jet as a b-jet, c-jet, u-jet, or τ-jet. This is achieved by designing algorithms that exploit information that is known from the SM about the associated hadrons.

Machine learning (ML) has had increasing application in the ATLAS experiment, one of them being HF tagging. Currently, the recommended ML b-tagging algorithm in ATLAS is known as DL1d, which is part of a two-level tagging approach that has been used in Run 2 and the baseline for Run 3 data-taking periods. It is part of the DL1 series, which is a multivariate algorithm. The two-level approach extracts information from the charged-particle tracks associated with the jet and uses ML algorithms to exploit decay properties of individual tracks and their relation to other tracks to reconstruct displaced vertices [53]. The outputs of the low-level taggers are fed into the high-level taggers to give us the jet flavour predictions as outputs. A detailed description of the different taggers in each level is provided in Chapter 5.1. The measurement of the b-jet cross-section made use of Run 2 data; the ATLAS recommended b-tagging tool for Run 2 is DL1r.

For a jet to be classified as a b-jet, there must be a b-hadron with $p_{\mathrm{T}} > 5$ GeV, must pass the 77% b-tagging efficiency working point, and must have a ghost associated with it. The identification of b-jets is p_{T}, which also in turn affects the b-tagging efficiency, so a scale factor is applied to MC events to account for differences in data and MC for the b-tagging accuracy at 77%. In the case of a c-jet, there was a hadron with a c-quark inside it; the hadron must have $p_{\mathrm{T}} > 5$ GeV. If the hadron does not satisfy the requirements for a c- or b-jet, it will be classified as a light quark jet.

The missing transverse momentum vector ($\vec{p_{\mathrm{T}}}^{miss}$) is reconstructed by applying strict conservation laws. It is defined as the negative vectorial sum of all the transverse momenta of all the detected and reconstructed final state objects in a collision. There is also a soft term added [54], which is the tracks linked to a primary vertex but have not been associated to a final state object. The magnitude of the $\vec{p_{\mathrm{T}}}^{miss}$ vector is referred to as E_T^{miss}. It is worth noting there are some challenges with reconstructing E_T^{miss} at particle level [54]; at detector level it includes mismeasurement of particles and detector effects which form part of unaccounted momentum, but this is not reconstructed at particle level.

Chapter 3

Exploration of Bottom-quark-philic Semi-visible Jets

The intention of this phenomenology study is to explore a DM signature that would arise if DM were assumed to be a particle that strongly interacts. In this proposed scenario, new signatures such as semi-visible jets (SVJs) [55] would be surveyed at colliders. This study focuses on semi-visible jet production with only bottom quarks in the final state. As previously mentioned, learning about the various elements that make up the SM, there are numerous unanswered questions around the existence of DM. With no evidence of DM thus far, a new signature is proposed. In this Chapter, a thorough outline of the building blocks of dark matter and the direct and indirect searches done to date. Finally, it explores several proposed models that have been discussed already. Later, a strong coupled dark sector is introduced as an unexplored signature, which is the main motivation behind its exploration, bringing us into the motivation for the signature, and finally, a series of discriminating variables are suggested as part of the search strategy.

3.1 Dark Matter

3.1.1 What Do We Know Thus Far About Dark Matter?

The origins of DM are a collection of precise astronomical inspections [56] that led to information known today, that close to 25% of the mass of the universe is non-luminous. From a classical perspective, this statement can be mathematically proven by closely analysing Zwicky's paper [56], which refines the velocity and density profile of DM [56]. Up till now, a greater part of what is known about dark matter is what it can not be, which has assisted in narrowing down the search. The lesser part of what is known is what form dark matter can take; this is linked to macroscopic features which give us details about particle properties. From an astronomy perspective, there is important information that is known, namely four things: (1) the quantity and distribution of dark matter, (2) how it interacts with baryonic matter, (3) dark matter is cold and (4) how long it has been there, which informs us about its stability.

In the earlier stages of when DM was postulated, it was believed that DM seeds were just modified gravity [56]; this quickly became apparent that it is impossible. DM seeds refer to the small fluctuations in the density of matter during the early universe that led to the formation of DM. Considering the amount of DM estimated to exist in the universe, there must have been a timely formation of structure [56]. If it is assumed that the early universe was a gravitational potential well, for structures to form, DM seeds must be electrically neutral to some extent [56] and non-relativistic. This prevents any form of seeds from free streaming out of the potential wells [56]. The detailed reasoning as to why modified gravity theories fall through can be found in [57], which uses baryonic density fluctuations to disprove the theory.

When it comes to theorising the particle structure of DM, there are two approaches: SM and non-SM. The non-SM approach explores possibly extending the SM to include DM candidates, whilst the SM approach proposes existing stable baryons, such as protons and neutrons, as DM candidates. Taking into consideration the precise observations that astronomy has provided us with, on cosmological scales one can already say that baryons do not work as DM candidates [56]. The simple reason behind this is that Big Bang Nucleosynthesis (BBN) sets constraints on the density of baryons in the early universe. Massive Compact Halo Objects (MACHOs) propose dark baryons as DM candidates. The searching for MACHOs entails the surveillance of Large Magellanic Clouds (LMC) with the goal of observing small magnification termed microlensing. Microlensing was ruled out when the results obtained from OGLE-III ([58]) and EROS-2 ([59]).

Another DM candidate that could be made of ordinary matter is primordial black holes (PBHs). PBHs are uncomplicated and easily formed, with no fixed predictions on their mass [56], which makes them completely model dependent. Constraints on PBHs can be derived based on specific scenarios. In ideal scenarios PBHs can be considered subdominant components of DM. The results obtained from PBHs are inconclusive, which is valid grounds for looking into extending the SM.

One of the downfalls when it comes to the notion of extending the SM is how intensive the model-building process is. Ensuring that the model is consistent with existing constraints and well-motivated theoretically whilst having limited literature is proven difficult. Whereas in the SM approach, there exist centuries worth of experiments, results and knowledge which tremendously assist in refining the new models. With that being said, over the past few years, many BSM particle models for DM have been developed.

3.1.2 Characteristics of a Dark Matter Model

When building a new model, the first thing one wants to do is ensure that there is a well-defined purpose for the model and the problem it's aiming to solve. The second important thing is establishing any existing boundaries or constraints for the model. For DM, it is of importance to know what the constraints are for its mass, interactions and relic density. Relic density is known to be the density of particles once they are in chemical and thermal equilibrium. In regard to mass, there are no pre-existing constraints that have been set on what the mass of the DM particle should be. Interactions explores how DM particles are predicted to behave and react with each other and with other forms of matter. Yet again, there are no strict constraints on the interaction properties. Lastly, relic density, which has been defined above, has one of the well-known constraints of DM. The reason for this is that astronomy has afforded us good approximations of the total density and velocity of DM at cosmological scales. This means that it must be ensured that the model being developed is consistent with the existing boundaries and values on abundance and distribution.

Closely looking at the zeroth-order lessons from cosmology leads to understanding DM during the early universe. By doing so, it introduces a particle model called weakly interacting massive particles (WIMPs). In the name itself, one can deduce that these particles interact through weak forces and are weakly coupled to the standard model. WIMPs are well-motivated when explored through the lens of thermal relics. Thermal relics are the particle species that were thermally created early on, when the universe was at a very high temperature and in chemical and thermal equilibrium. The number density of the particle species would be roughly equal to the number density of photons. As the universe cooled, the number of particles and photons would decrease together as long as the temperature remained higher than the mass of the particle species. When the temperature finally dropped below the particle mass, the creation of particles would require being on the tail of the thermal distribution. Metaphorically, it becomes a crossroad where statistical physics, general relativity, nuclear and particle physics meet. When the idea of cold relics was explored, it was well understood that it would mean that the mass of the particle would be greater than the freeze-out temperature [56]. The freeze-out temperature refers to a point in the early universe where the rate of particle annihilations drops below the rate of expansion of the universe [56]. It is an important component in understanding how WIMPs could have been created in abundance in the early universe through thermal processes. This also implies that the particle species will decouple in a non-relativistic regime and stop being in chemical equilibrium. The freeze-out temperature ensures that the number density remains constant whilst preserving the relic abundance that is observed today. At a temperature greater than the freeze-out temperature, these objects of interest redshift their number density away [56]. Calculating the correct relic abundance for WIMPs in a cold relic gave a value of 0.2; this value matched the expected relic abundance, which is how the term WIMP miracle came about. Cosmologists spent some time exploring what the mass limit would be on a particle like a WIMP, and the limit for DM candidates in a cold relic is 10 GeV [60]. From attempting to understand DM from a statistical physics point of view, one can get a better understanding of its behaviour and properties in the early universe. The important hypothesis one can carry over into the next few steps of building a particle model for DM is that DM production only took place in processes that were in thermal equilibrium and it has a zero chemical potential. With this being said, it is worth noting that there exist some theoretical frameworks where DM is produced non-thermally and where DM is not its own antiparticle [56].

3.1.3 Direct Detection of Dark Matter particles

Building on what has been discussed, there is still no information regarding what the particle flux, the particle's mass or what the expected kinematics would be in a WIMPs scenario. Using the knowledge on Lee-Weinberg's limit [60] and information on DM production in the early universe, the question that comes to mind is, how can one go about detecting WIMPs? It has been briefly mentioned that these massive particles would interact through weak forces, and in the SM the other particles that are known to weakly interact with matter are neutrinos. Neutrinos were discovered experimentally in 1956 [56], but they were anticipated to exist a decade before that. To quote Wolfgang Pauli, who said that he had done a terrible thing, he has postulated a particle that cannot be detected [61], in which he was referring to the existence of neutrinos. At that time, scientists were having trouble figuring out how they would go about detecting these particles. In Ref. [56], the author proposes using the knowledge and lessons attained in the discovery of neutrinos and applying it to directly detecting WIMPs. The first conceptualisation of direct detection of DM can be found in Ref. [62]. Over the years it has been developed into a concept called neutrino floor. It will be explored in detail in later sections, but particle colliders are not sensitive to SM neutrinos because they interact weakly with matter. The concept of a neutrino floor from a theoretical perspective is the lower limit in direct detection of WIMP-like DM models; it speculates that DM signals at lower limits resemble background neutrinos. It maps out all the possible sources of background neutrinos, estimates the expected particle flux in each process and attempts to improve the modelling of the background. The end goal is that if one were looking for the cross-section value for DM-nucleon interaction for a specific mass, the DM event rate would be equivalent to the neutrino background event rate. The author shares that he believes that reaching neutrino floor sensitivity will be an exciting point in particle physics [56].

With all that being said, it is worth noting that detecting particles of this nature is not easy. Especially since the goal is to eventually directly detect WIMP-like DM particles. If a case where DM scatters off a nucleus in an inelastic process working in the non-relativistic regime is being looked at, there are some values, such as the scattering cross-section between dark matter and nuclei and energy transfer to the nucleus, which give insights into the properties of the dark matter particle that can be obtained. This scenario was theoretically explored in depth in [63] from a model-independent approach. If a DM particle χ collides with an SM particle X, particle χ' is expected, which is the final state DM particle after the scattering and recoil of SM particle Y; one can attain the value for the threshold energy and maximum velocity of the recoil particle. By doing this, the energy deposit of χ' is expected to be in the KeV range [56]. From an instrumental perspective, a detector that can detect a particle massive enough with an energy scale of KeV, down to the fraction, is needed. There needs to be enough events and a large enough signal-to-noise ratio to obtain statistical significance. In Ref [56], an example with hypothetical values was made to estimate the rate of scattering in order to get a better understanding of whether this is something even detectable. They assumed a collision for 1 kg detector mass [56] with nuclei with radius R and the number of target nuclei per kg being 6.0×10^{26}, the event rate was estimated to be 0.06 events per kg per day [56]. This just means that an advanced detector placed in a strategic location is needed. It definitely raises questions if this is feasible and if looking into indirect detection would bring us closer to detection of DM.

3.1.4 Indirect Direction of Dark Matter

The notion of indirect detection is looking for by-products and not for the DM particle itself. That would entail looking for SM particles under the assumption that its origin can be traced back to DM. The two key processes in indirect searches are (1) pair annihilation and (2) decay of DM. Pair annihilation is when two DM particles annihilate into an SM particle, and decay of DM is when a single DM particle decays into an SM particle. There are conditions that one must be aware of as the focus is on these processes, the main one being that a non-vanishing coupling of the SM and DM sectors has to be presumed. To be specific, in order to study the decay of DM particles, the assumption that has to be taken is that DM can not be completely stable.

In the case of pair annihilation, it is worth noting that it can be suppressed if coupling of DM to SM is weak and if DM is not its own antiparticle. Other processes outside of pair annihilation and the decay of DM exist, but they are less explored and developed. As a result of (1) and (2) being refined and well-studied over the years, there is a promising prospect that is well-motivated for detecting non-gravitational signatures. Ultimately, indirect detection of these by-products is well-motivated under the assumption that DM is a thermal relic with the detection rate of these SM particles linked to DM origins. If this is the case, there might be some discrepancies in the annihilation rate today and the estimation made during chemical decoupling, but this is expected [56].

Indirect detection is made up of three important constituents [56]: production rates, energy scale, and annihilation products. These three constituents can be thought to be three main ingredients for indirect detection. When referring to production rates, it is the rate at which the SM particles are produced which will be messengers. It is important not to confuse it with the mediator, which acts as a portal between the hidden and visible sectors. Production rates depend on (1) the number of DM particles or particle pairs in regard to annihilation in a volume at that time [56], (2) the pair annihilation rates or decay rates, and lastly [56], (3) the number of SM particles per annihilation/decay [56]. The second constituent is energy scale; this relates to the expected energy range of the SM messenger. For direct detection of WIMPs, it was found that the expected energy deposition would be in the KeV energy range. The last ingredient is the particles produced after annihilation, which are completely model-dependent.

As mentioned above, the final state of any annihilation process is purely model-dependent. One can explore the model in which a pair of DM particles annihilate into lighter fermions with specific conditions on mass. In Ref. [56], the concept of DM particles being a Majorana fermion is explored, with $l = 0$ angular momentum state in which the expected final state would be $b\bar{b}$ and $\tau^+\tau^-$ [56]. Depending on the specific constraints on mass, the $\tau^+\tau^-$ final state can also be boosted, which means that the overall events will have a high transverse momentum. Another model for a final state of an annihilation process is if the DM particle is postulated to be a wino-like, the most dominant final state would be W^+W^- or even ZZ [64]. A wino-like particle is the lightest supersymmetry particle. This also highlights an overlap since the theory around supersymmetry is a signature searched for in particle colliders. To briefly explain, supersymmetry (SUSY) [65, 66, 67] is a BSM space-time theory that predicts fermionic partners for bosons and bosonic partners for fermions. It is evident that the particles that needed to be detected solely depend on the model of the DM particle.

On the other hand, in regard to the decay of DM particles, it has been mentioned that in order for this process to occur, DM particles would have to be unstable. With an adequately longer lifetime, there is a greater chance of being able to detect the decay products. This notion is well-motivated on the GUT scale and Planck scale [68]. The Planck scale refers to an even higher energy scale (around 10^{19} GeV), associated with quantum gravity and the unification of all fundamental forces, including gravity. While the GUT scale refers to a high-energy scale around 10^{15} to 10^{16} GeV, where the three fundamental forces of the SM are believed to unify according to Grand Unified Theory (GUT). The challenge that can be anticipated is that, with the decay of DM particles, it implies that DM candidates can be short-lived.

Some of the annihilation and decay products that have been searched for are charged cosmic rays, photons, and neutrinos [56]. In a scenario where DM annihilates into charged cosmic rays, it is anticipated to be a sub-dominant contribution to the decay products [56]. The challenge with charged cosmic rays is that they are not reliable messengers. As they travel to detectors, they lose energy and information relating to their origin. Since they are charged particles, they can be deflected by magnetic fields, altering their path. This means that extra care needs to be taken when modelling the transport of cosmic rays, taking into consideration their convection, diffusion, re-acceleration and fragmentation. Its initial physical condition also plays an important role in modelling the trajectory of these charged particles. There are two possible initial physical conditions. The first one is 'no diffusion, which assumes that the energy loss scale is much smaller in comparison with the diffuse timescale. This would give a steady-state and asymptotic solution. The second one is relevant to violent astrophysical sources such as galactic pulsars. In a burst-like injection from a point source, the solution to modelling would be a spherically symmetric solution that neglects energy loss.

Whereas, for neutrinos, they are perceived to be more reliable messengers because they are neutral; it means that they do not experience any form of deflection or energy loss as they travel to detectors. Neutrinos are the only particles produced during DM annihilation that can emerge from the centre of celestial bodies without losing energy [56]. The challenge is that, since neutrinos interact via weak force, they are difficult to detect, and the only two sources that have been observed thus far are the Sun and SN1987A [56].

In the case of photons, DM annihilates and produces a valuable multi-wavelength emission spectrum with expected secondary emission peaks [69]. The emission spectrum includes both prompt photons, which are from photon decay of neutral pions and internal bremsstrahlung, as well as secondary photons which originate from inverse Compton and synchrotron radiation. Assuming that the DM particles are WIMPs, they would produce photons in the gamma-ray energy range and beyond, with signal sources potentially being dwarf spheroidal galaxies, local Milky-like galaxies, local clusters, and galactic centres [56]. The challenge that is faced is there are not enough photon counts, which affects the statistical significance.

3.1.5 Searching for Dark Matter with Particle Colliders

The advantage of looking for DM using colliders is that one probes scenarios where DM particles are produced alongside SM particles and jets. Probing DM in colliders means a wide range of DM models with similar final states can be explored. The important aspect when it comes to looking for DM in colliders is the concept of missing transverse momentum. It has been established that DM weakly interacts with ordinary matter; this means that it will go undetected in the colliders. The other particles produced during the collision alongside DM particles are completely model-dependent.

One of the approaches that exist when exploring DM at colliders is the top-down and bottom-up approach. The top-down approach is based on a developed theory in which a distinct signature and signal are searched for in the colliders. Signal is the final state that is being searched for, whilst background would be other processes, which are not of interest, that would result in a similar process. An example of a top-down approach would be searching for SUSY. There are a few SUSY signatures, but in the search for DM, the signature that would be searched for is a final state with high missing transverse momentum, which is a result of the lightest supersymmetric particle (LSP), which is weakly interacting, and one or more SM particles, such as leptons or b-hadrons. The bottom-up approach is anticipating the behaviour and reaction between DM and SM particles with contact interactions. It is signature driven. An example of this would be using simplified models. The concept of a simplified DM model means there must exist a DM particle that is stable and weakly interacting so that it goes undetected in colliders; there would also need to be a mediator particle that acts as a portal between DM and SM particles. An outline of searches for DM in colliders can be found [70]. An example of this would be the case of WIMPs. DM particles will not be detected in the colliders since it is hypothesised that they are weakly coupled to the SM. When reconstructing the final state, there will be missing transverse momentum. Mono-X signatures in colliders arise in the WIMPs scenario. A mono-X signature has a large E_T^{miss} because the SM particle X recoils against an invisible DM particle. The signature's SM particle X could be a photon, top quark, W or Z vector boson, or even a Higgs boson. The stability of the DM particles, which results in E_T^{miss}, is a crucial assumption in the WIMPs scenario. An overview of WIMP searches is also provided in Ref. [70].

To optimise the search for these specific signatures in colliders, specific requirements are applied on kinematics, such as jet multiplicity, transverse momentum of the jets and other SM particles such as electrons and identifying background processes. For most DM searches, the dominant background processes are any processes that produce neutrinos. Neutrinos, just like DM, do not interact with ordinary matter, resulting in missing transverse momentum.

3.2 What are Semi-visible Jets?

Despite the numerous direct and indirect searches conducted over the past ten years in all areas of physics, DM has not yet been found. Detecting DM in colliders has been one of the goals of the LHC [71]. The only well-known characteristic of dark matter is that it hardly interacts with regular matter, which means it will not be detected in colliders.

It is necessary to explain the dark sector before defining the semi-visible jet (SVJ) signature. According to Ref. [72], the dark sector is a hypothetical component of the universe that differs significantly from the SM. The dark sector, in contrast to the visible sector, known as the SM, presents particles that can't be seen. It was hypothesised that DM would interact weakly with SM particles and fundamental forces if it existed. Two parameters, R_{inv} and α_{dark}, can be examined to comprehend the dynamics in the dark sector. R_{inv} is the name for the ratio of the stable dark hadrons generated to the total number of hadrons simulated in the event. The R_{inv} stands for $R_{invisible}$, referring to the dark hadrons not being part of the visible sector, which is the SM. In essence, it captures the dark sector's transition from stable to unstable hadrons [55]. The R_{inv} approaches zero [55] if dark hadrons are present in an event and they decay entirely to SM particles. In that case, the collider signature would be just a QCD parton shower without any E_T^{miss}. Similarly, the value of R_{inv} would approach 1 [55] if the dark hadron did not decay into visible particles. In this scenario, the signature would be E_T^{miss} in the direction of visible jets. Ref. [73] provides the theory and motivation for dark strong coupling parameters, elaborating on the appropriate parameter values that should be considered when creating simulated events. Input parameters for confining scales based on gauge theories and characteristic mass scales for dark hadrons are required by the program that simulates signal and background events. It is crucial to determine the best value. The number of dark jets, the parton shower's approximate transverse momentum, and, eventually, the missing transverse momentum and total energy of the simulated SM particles are all impacted by the strength of these parameters. High object multiplicity, collider objects like leptonic, photon or hadronic jets, and deranged vertices are the event topologies that can be searched for in colliders and that emerge experimentally from strongly coupled particles.

It has been mentioned what topologies arise when R_{inv} takes the value of 1 and 0; in the case of intermediate values, it is expected to have both dark and visible hadrons. Semi-visible jets (SVJ) [55] are a result of R_{inv} taking on intermediate values. This means that the final state will consist of an invisible and visible component. The invisible components, which would be the DM particles, would be produced in a QCD-like parton shower [71] as mentioned above through a mediator. The mediator is a particle that acts as a portal between visible and dark sectors; it mediates the creation of dark quarks, which results in stable and unstable dark hadrons. The decay to SM particles is visualised in Figure 3.1 and can occur in three different ways: contact operator, s-channel, and t-channel. The mediator particle is classified as a Z' boson [55] in the s-channel and as a bi-fundamental mediator [74] in the t-channel. The CMS collaboration [75] has provided a result for the resonant channel production, successfully setting the lower limit on mediator masses up to 4 TeV. The WITS ATLAS group reported the first search for non-resonant production of SVJs in Ref. [74], establishing the first upper limits for the strongly coupled dark sector for a bi-fundamental mediator mass up to 2.7 TeV.

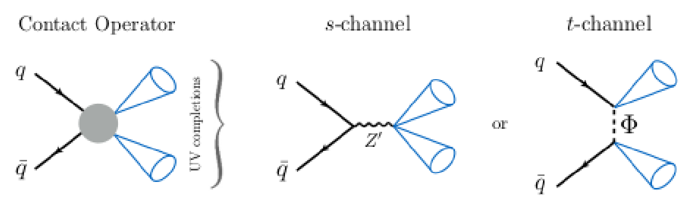

Figure 3.1: Various portal through which DM particles may decay into SM particles, taken from [55]

In colliders, where jets could be interspersed with dark matter particles, SVJs are among the most fascinating indicators of strongly interacting massive particles (SIMPs). SVJ signatures have mostly been unexplored because these events are usually dismissed as indications of jet mismeasurement. With E_T^{miss} aligned along one of the jets, the features should be back-to-back SVJ. Figure 3.2 illustrates the collider signature for WIMPs and shows the expected alignment of the E_T^{miss} with respect to the visible jets.

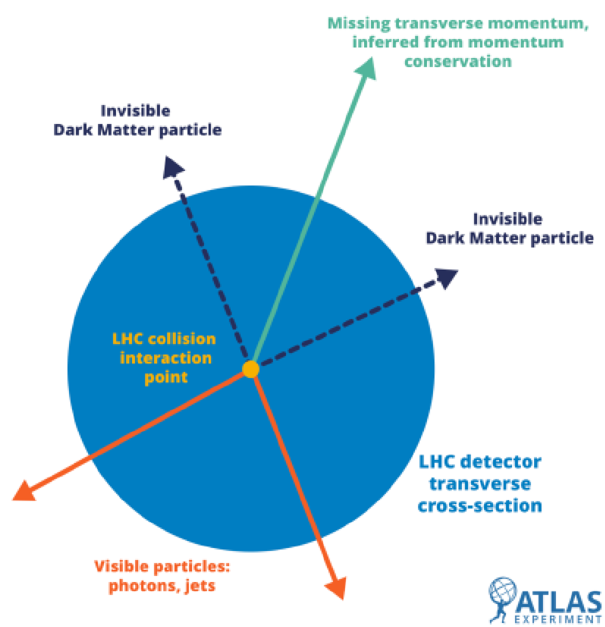

Figure 3.2: The illustration depicts the event topology, indicating where the direction of the missing transverse momentum balances the jets, adapted from [76]. This is a collider signature for WIMPs

Now that SVJs have been explained, it is important to note that the existence of a WIMPs scenario does not rule out the possibility of a SIMPs scenario. Proposing DM to be strongly interacting massive particles gives rise to collider topologies such as SVJs. The dark sector does not only give rise to SVJ but also has many different signatures, such as dark jets and emerging jets. There is a lot of new physics pertaining to undiscovered quantum fields and the particles that go along with them in this completely unexplored world. Having said that, the characteristics to be aware of in this instance are that dark showers would occur when the dark sector is strongly coupled to the SM. The intriguing aspect of this story is that it raises the possibility that the DM has existed inside hadronic jets.

3.3 Motivation for SVJ-b Signature

This is the first search for SVJs with b-quarks (SVJ-b), which raises the level of complexity. The motivation behind proposing this signature with b-quarks is because the combination of displaced vertex and high mass gives b-hadrons a distinctive signature in experimental collisions. As depicted in Figure 3.3, the primary vertex is where the hard scattering of two protons occurs, and the secondary vertex, which is the displaced vertex, is shown. A more in-depth discussion about what is b-tagging and how it is done in ATLAS, alongside the different algorithms used, is presented in Chapter 5.1. In previous searches of SVJs, it was inclusive of jets containing all hadrons. In this specific signature SVJ-b, it requires the jets to contain b-hadrons, to not only greatly lower the leading multijet background but also to increase the discriminatory power to detect them. This scenario provides a novel probe of this completely undiscovered signature and has been inspired by theoretical papers ([77], [78], [79], [80]).

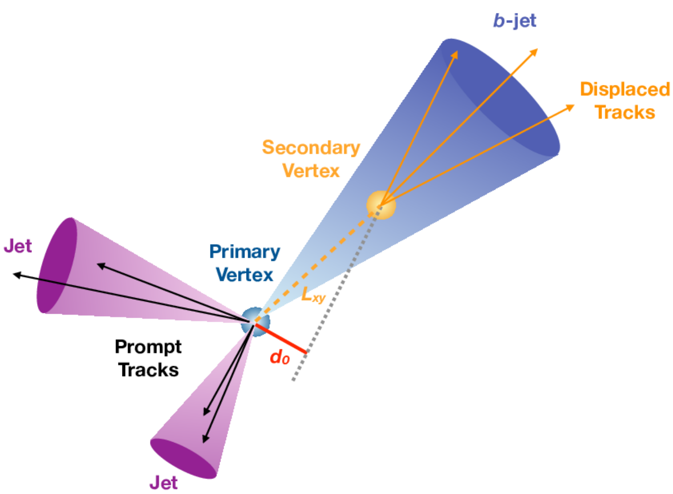

Figure 3.3: b-hadrons are known to have a distinctive characteristic of having a longer lifetime. This means that they travel resulting in displaced vertex as shown in the schematic diagram, adapted from [81]

3.4 Simulating SVJ-b Signal and Background

When it comes to simulating this signature, since it is a continuation from previous work done by the WITS ATLAS group, the modelling of the signal and background is based on Ref. [74]. In Figure 3.4, the Feynman diagram of the signature simulated is shown. The signal was generated using MadGraph5_aMC@NLO to produce the matrix element. MadGraph5_aMC@NLO with two extra partons at leading order with $\sqrt{s} = 13$ TeV centre of mass was generated using the t-channel and simplified dark matter model [82]. When it comes to hadronization and the patron shower, it was simulated using `PYTHIA` [11, 12]. For the parton distribution function, NNPDF3.0LO [83] was used.

The Hidden Valley (HV) module [84, 85] in `PYTHIA` 8 was utilised to (1) simulate the portal between

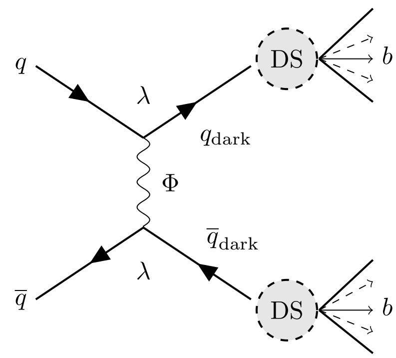

Figure 3.4: A feynman diagram illustrating the production of semi-visible jets in t-channel. This figure is adapted from Ref [74]. The symbol, Φ denotes the mediator that acts as a portal between the visible and dark sector. DS stands for dark shower, which would decay into SM particles. In this particular signature, a final state with b-jets is explored. One of the free parameters in this model, λ which is the coupling strength between the SM and the dark sector and the other free parameter is R_{inv} which governs the amount of dark unstable hadrons that would decay into stable visible SM b-hadrons.

the dark sector and visible sector and (2) set the appropriate value for the parameter of the dark coupling α_{dark} based on Ref. [73]. It is designed to take care of any sector decoupled from the SM. The interaction between SM and HV is allowed due to TeV-scale operators. To be specific, the HV module was used to shower the matrix-level events and produce dark hadrons. The number of flavours in HV was set to two, creating two dark mesons π_d and ρ_d. The vector dark meson ρ_d would promptly decay into a pair of pseudo-scalar dark mesons π_d. In the matrix element, the flavour-diagonal π_d and ρ_d meson masses were set to 20 GeV and 40 GeV, and the off-diagonal π_d and ρ_d meson masses were set to 9.99 GeV and 19.99 GeV, motivated by Ref. [86, 73]. The π_d pair would then decay into SM quarks through a bi-fundamental mediator (Φ) observed in Figure 3.4. As mentioned, these values allow for heavy-flavour decay and helicity suppression. The helical suppression allows for the pseudo-scalar dark meson pair to decay into $b\bar{b}$ to be dominant, provided that certain mass conditions are satisfied. The mass of the dark hadron is set to be 10 GeV based on Ref. [55], and the mediator particle is set to be 3000 GeV, based on Ref. [74]. The MLM jet matching scheme was used, and the matching parameter was set to 100 GeV. There are two free parameters in the model, one of them being R_{inv}. It is defined to be the fraction of unstable dark mesons that decay into stable dark mesons; it is worth noting that different values of R_{inv} result in different topologies. The signal was produced using $R_{inv} = 0.33$ and $R_{inv} = 0.67$. The second free parameter in the model is the coupling constant between the SM and DM sectors. In this particular modelling of the signal, in the HV module, α_d was set to be running at one loop. The dark confinement scale λ_d discussed in [73] was set to 6.5 TeV. The dominating backgrounds were identified to be multijet and $t\bar{t}$, and they were generated using `PYTHIA`. The background processes were modelled to understand and characterise the underlying kinematic features. Rivet Toolkit [87] was used to analyse the results, and Fastjet package [51] was used for jet clustering.

3.5 Current Constraints on SVJ-b Signature

In the paper [74], it was the first time that the SVJ signature in the t-channel was explored in much detail. CMS has explored the signature in the s-channel and can be found in [75]. In the search for SVJ in the t-channel, a better understanding of the event characteristics was achieved, and masses up to 2.7 TeV were excluded. This time around the intention is to see if any existing search already excluded the SVJ-b signal. A thorough literature review yielded three ATLAS searches that had a similar final state to the SVJ-b at the time. These ATLAS searches are:

1. The search for supersymmetry in final states with missing transverse momentum with two or more b-jets [88]
2. The search for dark matter produced in association with a Standard Model Higgs boson decaying

into b-quarks [89]

3. Search for non-resonant production of semi-visible jets [74]

All the listed are ATLAS analyses that use full Run 2 data collected at the centre of mass $\sqrt{s} = 13$ TeV at a luminosity of 139 fb^{-1}. The data was all available on HEPDATA [90]. To test if any of the three analyses set constraints on the SVJ-b signature, the signal was passed through the Rivet analyses, which were either public or soon to be public. The recasting was done at the particle level. Recasting is when the signal, in this case SVJ-b, is passed through existing measurements and searches that have a similar final state to the signal. This is done in order to determine if the signal is excluded or not. For the first search, the validation of the rivet analysis was performed both by comparing distributions after 0-lepton pre-selection and yields in the specific signal regions (SRs); for the second and third analyses, the yields for the benchmark signal from the written rivet analysis were compared to yields reported by ATLAS for 2-$b$jet and 3-$b$jet SRs and 9 bin SR for the non-resonant production of SVJ.

So in order to see the sensitivity of the SVJ-b signal, the search was implemented in the Rivet MC Analysis toolkit [87]. The paper presents a search for pair-produced gluinos decaying via top or bottom squarks in events with b-jets, high missing transverse momentum and potentially additional jets or isolated leptons. The rivet analysis was not publicly available at the time, so an analysis was written based on the paper. SUSY predicts new bosonic partners for fermions of the Standard Model and new fermionic partners for its bosons. Gluinos, new fermionic partners for the gluons, are expected to be pair-produced with a high cross-section at the LHC. The paper focuses on gluinos decaying into the lightest SUSY particles, known as the neutralinos. $\tilde{\chi}_1^0$. In the paper, three simplified signal benchmark models were considered: (1) $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$, referred to as the Gtt model; (2) $\tilde{g} \to b\bar{b}\tilde{\chi}_1^0$ referred to as the Gbb model and (3) $\tilde{g} \to t\bar{b}\tilde{\chi}_1^-$ and $\tilde{g} \to b\bar{t}\tilde{\chi}_1^+$ known as the Gtb.

Out of all three signal models mentioned, the main focus was on recreating the kinematic distributions for the Gbb signal model. The reason behind this is the Gbb signal model probes a similar final state to SVJ-b. Figure 3.5 shows the decay processes for the Gbb benchmark processes.

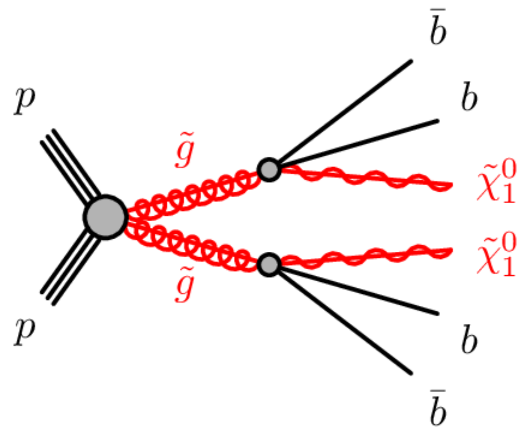

Figure 3.5: Schematic diagram for the decay process for Gbb benchmark model. Adapted from [88]

For event reconstruction, a primary vertex with at least two tracks with $p_{\mathrm{T}} > 500$ MeV was required. Jets were reconstructed using anti-k_t algorithm [50] with $R = 0.4$. The jets including b-tagged jets were required to have $p_{\mathrm{T}} > 30$ GeV, $|\eta| < 2.8$ and MV2c10 algorithm was used for differentiating b-jets from other jets. The electrons and muons were required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. Only events with $E_T^{miss} > 200$ GeV were considered for the analysis.

For the analysis strategy, the signal regions were defined and explored using two methods: cut-and-count and neural networks analysis. Both methods used five discriminating variables to achieve maximum exclusionary power. This was achieved by rejecting SM background events and retaining as much signal as possible.

The first discriminating variable is inclusive mass effective, which is defined as:

$$m_{eff} = \sum_i p_{\mathrm{T}}^{jet_i} + \sum_i p_{\mathrm{T}}^{l_j} + E_T^{miss} \tag{3.1}$$

This variable informs us of the invariant mass of the final state, as it is expected for the final state to have E_T^{miss}, jets and possible leptons.

The second discriminating variable is transverse mass. This variable only applies to regions with at least one selected lepton, and its defined as:

$$m_T = \sqrt{2p_{\mathrm{T}}^l E_T^{miss} 1 - \cos[\delta\phi(\vec{p}_T^{miss}, \vec{p}_T^l)]} \tag{3.2}$$

The third variable is minimum transverse mass which is formulated using E_T^{miss} and 3 highest p_{T} b-tagged jets, and is calculated as follows:

$$m_{T,min}^{b-jets} = min_{i\leq 3}\left(\sqrt{2p_{\mathrm{T}}^{b-jet} E_T^{miss} 1 - \cos[\delta\phi(\vec{p}_T^{miss}, \vec{p}_T^{b-jet})]}\right) \tag{3.3}$$

Additionally, there is a total jet mass variable. This variable only takes into account the mass of large-R jets in events of a high p_{T} final state. It is a useful variable for the Gtt signal model, where boosted top quarks would decay into a single large-R jet. The function is defined as:

$$M_J^{\sum} = \sum_{i\leq 4} m_{J,i} \tag{3.4}$$

Lastly, this variable is only relevant in events with 0 signal region leptons. The function should find the minimum azimuthal angle between jet p_{T} and MET for 4 leading jets. The variable is calculated as:

$$\Delta\phi_{min}^{4j} = min_{i\leq 4}\left(|\phi_{jet} - \phi_{\vec{p}_T^{miss}}|\right) \tag{3.5}$$

The purpose behind this variable is to provide information on the relationship between the direction of E_T^{miss} and the direction of leading jets.

The generated plots were made using loose preselection requirements. All events were expected to have at least 4 selected jets; 3 of the selected jets had to be b-tagged jets. Additionally, all events considered had E_T^{miss} greater than 200 GeV, for events with 0 leptons in the signal region were additionally required to have $\delta\phi_{min}^{4j} > 0.4$. After the Rivet analysis was written, taking into consideration all the preselection requirements mentioned above and running the same signal that was used in Ref. [88], a comparison of their results was made to the results that were obtained. As mentioned, the main focus was on the Gbb signal model since it is the most relevant to the final state that is being probed in ATLAS. In Figures 3.6, 3.7, 3.8, 3.9, 3.10, and 3.11, when comparing the results obtained from [88] and the histograms plotted using Rivet, both were based on the preselection requirements from [88]; the distributions exhibited similar trends. In Figure 3.6, there is a decreasing trend with the plot that was reproduced, peaking at 5 selected jets; this was similar to the distribution found in Ref. [88]. Figure 3.7 displays a staircase-like decrease for the number of b-tagged jets in each event, with most events having 3 b-tagged jets. In Figure 3.8, it was the least well-modelled by the analysis, but this is expected, as producing E_T^{miss} at the particle level can be challenging. In Figure 3.9, on the other hand, it depicts the variable m_{eff}, which was previously mentioned before as the invariant mass scale; it displayed a similar trend to the m_{eff} distribution found in [88]. The only difference is a small gap in m_{eff} in the reproduced analysis close to 1000 GeV. Lastly, Figure 3.10 and Figure 3.11 were well-modelled by the analysis, reproducing similar trends to distributions found in Ref. [88], the difference being the scaling on the y-axis.

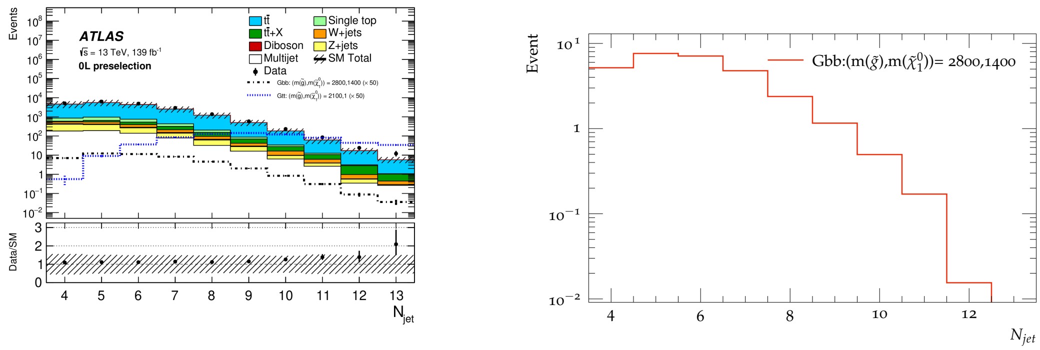

Figure 3.6: The distribution (a) on the left-hand side is taken from [88], showing the number selected of jets in each event. The histogram (b) on the right-hand side shows the number of events plotted against the number of selected jets in each event. Comparing the results from (a) and (b) after passing the signal Gbb: (m($\tilde{g}$),m($\tilde{\chi}_1^0$)= 2800,1400) the distributions have the same decreasing trend. The histogram that was recreated peaks at 5 jets then proceeds to decrease.

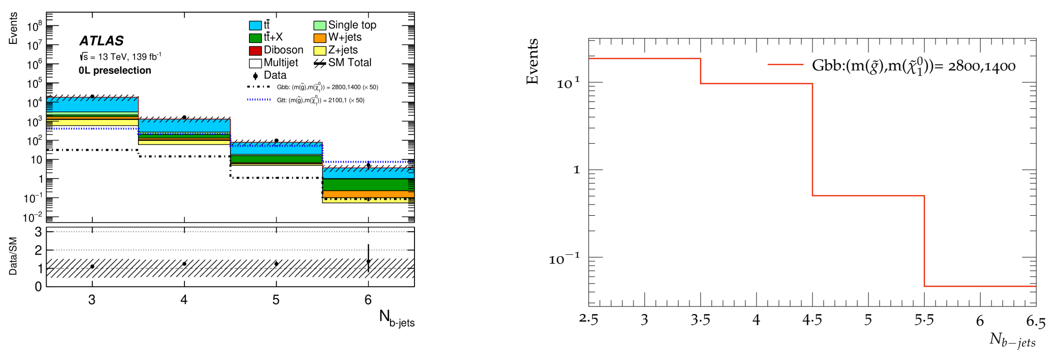

Figure 3.7: The distribution (a) on the left-hand side is taken from [88], showing the number of selected b-tagged jets in each event. The histogram (b) on the right-hand side shows the number of events plotted against the number of b-tagged jets. Comparing the results from (a) and (b) after passing the signal Gbb: (m($\tilde{g}$),m($\tilde{\chi}_1^0$)= 2800,1400) the distributions have the same decreasing trend.

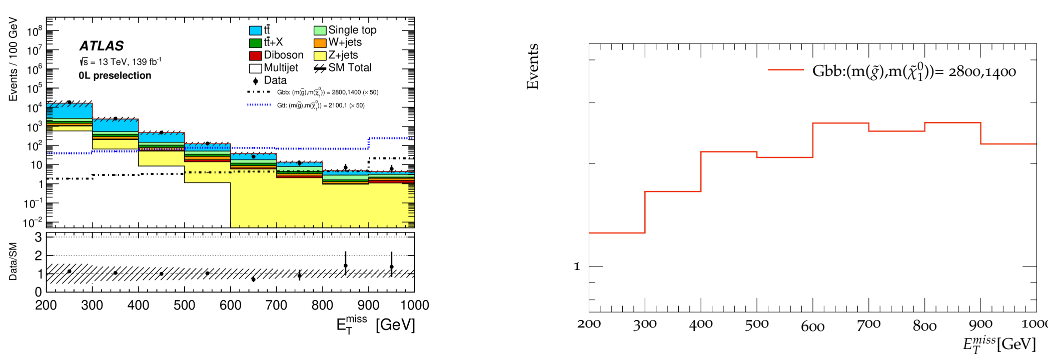

Figure 3.8: The distribution (a) on the left-hand side is taken from [88], displaying the E_T^{miss} measured in each event. The histogram (b) on the right-hand side shows the number of events plotted against E_T^{miss}. Comparing the results from (a) and (b) after passing the signal Gbb: (m($\tilde{g}$),m($\tilde{\chi}_1^0$)= 2800,1400) the distributions have the same decreasing trend.

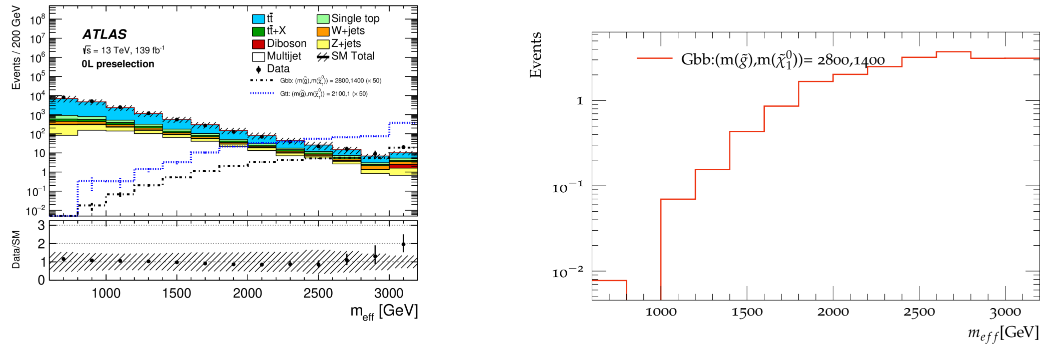

Figure 3.9: The distribution (a) on the left-hand side is taken from [88], displaying the measured m_{eff} in each event. The histogram (b) on the right-hand side shows the number of events plotted against the variable m_{eff}. Comparing the results from (a) and (b) after passing the signal Gbb: (m($\tilde{g}$),m($\tilde{\chi}_1^0$)= 2800,1400) the distributions have a similar gradual increasing trend.

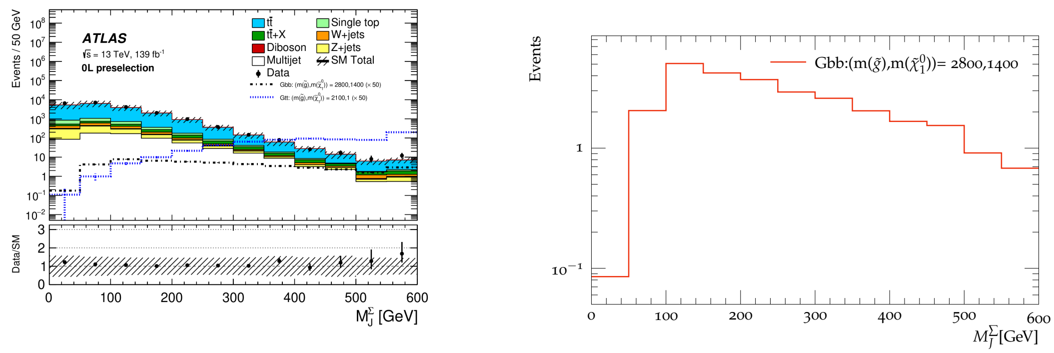

Figure 3.10: The distribution (a) on the left-hand side is taken from [88], showing the measured $M_J^{\sum}$ in each event. The histogram (b) on the right-hand side shows the number of events plotted against the variable $M_J^{\sum}$. Comparing the results from (a) and (b) after passing the signal Gbb: (m($\tilde{g}$), m($\tilde{\chi}_1^0$)= 2800,1400) the distributions have the same trend. Initially, there is an increasing trend that peaks at around 100-200 GeV, then the trend gradually decreases.

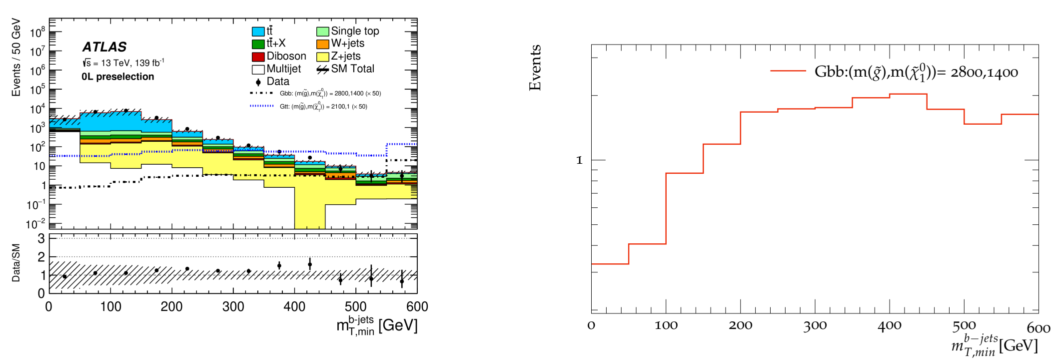

Figure 3.11: The distribution (a) on the left-hand side is taken from [88], showing the measured $m_{T,min}^{b-jets}$in each event. The histogram (b) on the right-hand side shows the number of events plotted against the variable $m_{T,min}^{b-jets}$. Comparing the results from (a) and (b) after passing the signal Gbb: (m($\tilde{g}$) ,m($\tilde{\chi}_1^0$) = 2800, 1400) the distributions have the same staggered increasing trend.

After obtaining these results, the next step was to pass the signal being probed through the analysis to determine if it is excluded by this ATLAS analysis. Since the main focus was on the Gbb signal region with 0 leptons, there were additional final event selection requirements. The SR was divided into three regions: Region B, Region M and Region C. Region B is classified to be boosted with a large invariant mass, including events with $m_{eff} > 2600$ GeV being a part of Region B. Region M was chosen to have a moderate invariant mass, with the requirement of $m_{eff} > 2000$ GeV. Region C contains compressed events with a small invariant mass and $m_{eff} > 1600$ GeV. All three regions required a $E_T^{miss} > 550$ GeV and $m_{T,min}^{b-jets} > 130$ GeV. Region M and Region C required the leading four jets to have $p_{\mathrm{T}} >$ 30 GeV, whilst Region B $p_{\mathrm{T}} > 65$ GeV. Adding in these new SR requirements and then passing the probed signal through the analysis, no signal events passed the SR selections. This means that this ATLAS analysis is not sensitive to the SVJ-b signal, concluding that it is not excluded.

An attempt to produce the SR yields found in Figure 3.1 and according to data, SR Gbb-B had 7 events, SR Gbb-M had 18 events, and SR Gbb-C had 32 events, as shown in Table 3.2. In the same Table 3.2, it shows the expected SR yields based on the analysis written. To possibly explain the reason behind not observing any SVJ-b signal yield, the Table 3.1 provides some insight. The initial expected yields were 80.9 and dropped to 56.6 after the common requirements were implemented. All the signal regions have the same E_T^{miss} requirement with varying m_{eff} values required; the motivation behind this has been mentioned. The reason why no signal events were observed in Table 3.2 was due to the common requirement because a requirement of 200 GeV was already implemented. Looking at the requirements for Gbb-C, there would be some events present since it has the lowest m_{eff} requirement and all SRs require at least 550 GeV. Overall, even though the distributions and yields are not identical, they are in the same order of magnitude, which is acceptable for reproducing a detector-level analysis at the particle level with high missing transverse momentum values.

	Selection	$m(\tilde{g})$= 2000 GeV, $m(\tilde{\chi}_1^0)$= 1000 GeV
		80.9
Common Requirements	$N_{\mathrm{lepton,base}} = 0$	80.0
	$\Delta\phi_{\mathrm{min}}^{4\mathrm{j}} \geq 0.4$	61.1
	$m_{\mathrm{T,min}}^{b\text{-jets}} \geq 130$	56.6
SR-Gbb-B	$E_{\mathrm{T}}^{\mathrm{miss}} \geq 550$	35.69
	$p_{\mathrm{T}}{}^{\mathrm{jet}} \geq 65$	35.69
	$m_{\mathrm{eff}} \geq 2600$	10.13
SR-Gbb-M	$E_{\mathrm{T}}^{\mathrm{miss}} \geq 550$	35.69
	$m_{\mathrm{eff}} \geq 2000$	28.30
SR-Gbb-C	$m_{\mathrm{eff}} \geq 1600$	35.69
	$E_{\mathrm{T}}^{\mathrm{miss}} \geq 550$	34.71

Table 3.1: Table showing the expected yields for signal region Gbb-B, Gbb-M and Gbb-C. The signal was generated using 30 000 events. The cutflow table was adapted from Ref [88]

	Benchmark signal		Data	SVJ-b signal
Selection	ATLAS yield	our yield	yield	yield
SR-B	10.13	15	7	0
SR-M	28.30	21	18	0
SR-C	34.71	22	32	0

Table 3.2: Table showing the number of entries and expected yields produced based on the analysis that was written and compiled based on Ref [88].

For the second analysis, the distributions in [89] were not reproduced as above but instead compared the yields obtained for this benchmark signal to yields reported by ATLAS for 2 b-jet and 3 b-jet SRs. In Section 3.1.5, what was discussed is the different ways in which dark matter is searched for in colliders. In this analysis, they search for dark matter using a mono-X signature, which in Section 3.1.5 was mentioned as a result of assuming that DM particles are WIMPs. The X in mono-X, as mentioned, stands for visible SM particles, and in this search, they explored the mono-Higgs signature. Assuming

2 b-tagged SR					
	Benchmark signal		Data	SVJ-b signal (R_{inv}= 0.33)	SVJ-b signal (R_{inv}= 0.67)
Selection	ATLAS yield	our yield	yield	yield	yield
$150 \leq E_{\text{T}}^{\text{miss}} < 200$ GeV	60	110	14259	16	12
$200 \leq E_{\text{T}}^{\text{miss}} < 350$ GeV	70	100	13724	40	30
$350 \leq E_{\text{T}}^{\text{miss}} < 500$ GeV	3.6	6	799	8	6
3 b-tagged SR					
$150 \leq E_{\text{T}}^{\text{miss}} < 200$ GeV	5.3	9	408	1.1	0.8
$200 \leq E_{\text{T}}^{\text{miss}} < 350$ GeV	18	7	658	6.8	2.6
$350 \leq E_{\text{T}}^{\text{miss}} < 500$ GeV	2.9	0.5	42	0.9	0.7

Table 3.3: Table showing yields in 2b and 3b resolved signal region for benchmark signal points, data from [95], and from our Rivet analysis and for SVJ-b signals.

the coupling of the Higgs boson to light quarks is suppressed, then the Higgs would be produced in the same mechanism that would produce the DM particle or through final-state radiation. The reason this is beneficial, in case of a discovery, would be to provide information about coupling between SM and DM. The search proposed an extension of the SM to include a two-Higgs-doublet model, also known as 2HDM [91], which predicts a production of five Higgs bosons after mixing using two simplified benchmark models: Z'-2HDM [92] and 2HDM+a [93, 94]. To check if the search set any constraints, the benchmark model with b-associated production was used, which is the 2HDM+a model. The Venn diagram for this signature can be seen in Figure 3.12. This benchmark model is the simplest renormalisable and gauge-invariant extension of a simplified mediator model [89].

To further explain the model, there is a pseudoscalar singlet that mediates the interaction between SM and DM candidates denoted as χ. The coupling between DM and SM is labelled y_χ, and the mass of the DM is labelled as m_χ; both these are free parameters in the model. The singlet mixes with pseudoscalar A from the Higgs doublet at a certain angle and results in pseudoscalar a. Using the 2HDM+a model is favourable because it can result in many experimental signatures, which means it can be used to exclude numerous regions. The regions of interest for the b-production of the 2HDM+a model were divided into two: 2 b-tagged SR and 3 b-tagged SR. It required events with E_T^{miss} between 150 and 500 GeV, no charged lepton, and at least two b-tagged jets. The sum of 2 jets is required to be greater than 100 GeV, and none of the jets had to be within 20°. The benchmark model used to compare the yield obtained using the SVJ-b signature was $(m_A, m_a) = 300$, 150 GeV. When comparing the ATLAS yields reported in the search with the yields obtained using the benchmark signal, it was within 100%, taking into account that it is without acceptance times efficiency. The numbers obtained are at the particle level, knowing that it can be difficult to reproduce E_T^{miss} at the particle level. When comparing data yield and yields from the SVJ-b signal with $R_{inv} = 0.33$ and $R_{inv} = 0.67$, both the yields reported on were smaller than the data of the signature, making them negligible. Overall, the results tabulated in 3.3 allow for us to conclude that this analysis does not exclude the SVJ-b signal.

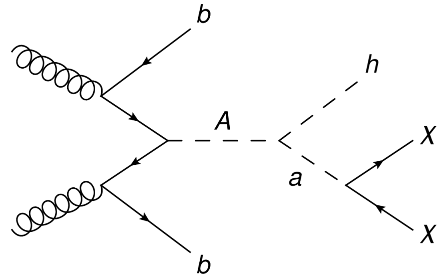

Figure 3.12: The feynmann diagram of the 2HDM+a signature. The diagram is taken from [89]

The third analysis forms part of the SVJ-b signature. The only difference being instead of looking for SVJ with all five quark flavours, SVJ-b deals with a final state with b-jets. The rivet routine of the analysis was publicly available; the SVJ-b signal was passed through the analyses to test the sensitivity. The SVJ required events with at least 2 jets, with a leading $p_{\text{T}} > 250$ GeV. One of the 2 jets required had to be within $\Delta\phi$ of 2 from the direction of the missing transverse momentum, and there had to be no charged leptons. The scalar sum p_{T} of all jets and E_T^{miss} of at least 600 GeV. The data from the

SVJ signal was divided into 9 bins; the results shown in Figure 3.13 showed the SVJ-b signal denoted as the signal with a mediator mass of 3000 GeV with $R_{inv} = 0.33$ and the signal with a mediator mass of 3000 GeV with $R_{inv} = 0.67$ is negligible when compared to SVJ. It doesn't constrain the SVJ-b signal, as the yields were much less than the data. The results obtained were justified and feasible. The cross-section of light-flavoured jets is greater than the cross-section of b-jets; the cross-section coupled with no strict requirement of b-jets meant that the SVJ analysis was not going to set constraints on the signal.

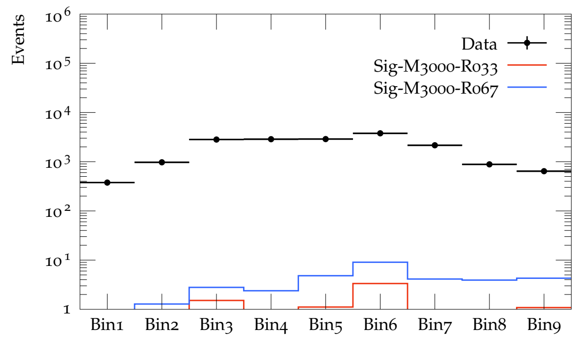

Figure 3.13: Histogram comparing the data yields of SVJ taken from [74] which the signal region was divided into 9 bins and the SVJ-b benchmark yields denoted by Sig-M3000-R033 which is the signal with mediator mass of 3000 GeV with $R_{inv} = 0.33$ and Sig-M3000-R067 which is signal with mediator mass of 3000 GeV with $R_{inv} = 0.67$.

3.6 Signal Reconstruction

The initial starting point for the signal reconstruction was taken from [74] since the signature that is being explored is similar. In [74], the E_T^{miss} trigger was used; therefore, a requirement of events to have $E_T^{miss} > 200$ GeV was implemented in the pre-selection and $E_T^{miss} > 600$ GeV requirement in the signal region. For jets, the anti-k_t algorithm was used with $R = 0.4$, with charged and neutral constituents used as inputs. There was a requirement of the leading jet p_{T} to be 250 GeV and other jets to at least have p_{T} of 30 GeV. It required at least 2 jets, and to be considered an SVJ candidate, at least one of the jets had to be $\Delta\Phi < 2$ in the direction of p_{T}^{miss}. Lastly, a harsh veto was applied on charged leptons of 7 GeV or more and events with 2 or more b-tagged jets to eliminate $t\bar{t}$ background. The aim was to reconstruct the signal for SVJ-b by implementing certain requirements inspired by the ATLAS non-resonant search for SVJs.

In the non-resonant search for SVJ, the 200 GeV E_T^{miss} pre-selection cut reduced the number of events in the final state significantly when doing it at the particle level. If SVJ-b were to be a search in the future, the $E_T^{miss} > 200$ GeV pre-selection requirement would not be an issue at the detector level, as events have higher missing transverse momentum due to detector effects and mismeasurements. At this current stage, removing this requirement would assist in increasing signal acceptance. With this being said, the final state that is desired would have b-jets and E_T^{miss}; it would be worthwhile to look into a new possible jet clustering algorithm that would cluster as many jets as possible.

Using the information from a study of jet substructure of SVJ signature [96], it was noted that the jets tend to spread out and have gaps between them. As part of the signal reconstruction, various jet clustering algorithms were used and compared against each other. To compare the different jet clustering techniques, an event selection closely related to the SVJ event requirement was implemented. The pre-selection required at least 2 b-tagged jets, where one of the jets had to be within $\Delta\Phi < 2.0$ of the p_{T}^{miss} direction with no restriction on the leading jet p_{T}, and the E_T^{miss} 200 GeV pre-selection was removed to improve statistics.

In ATLAS, the anti-k_t jet clustering algorithm with $R = 0.4$ is commonly used. For the comparison anti-k_t jets with p_{T} of at least 30 GeV were used, which is possible using the E_T^{miss} trigger but not for single jet triggers which have high p_{T} for their leading jet. Large radius jets [97, 98], also known as large-R jets, also used the anti-k_t fixed cone radius algorithm, but instead of using a radius of 0.4, it uses $R = 1.0$, which can be useful in boosted topologies. To use large-R jets, two b-tagged jets with a p_{T} of at least 200 GeV would have to be required. The reason for this specific pT requirement is that it is commonly used in boosted topologies, which are events with high transverse momentum. Another option mentioned is the reclustered (RC) jet [99] and dynamic radius (DR) jets clustering algorithm [100]. Reclustered jets make use of already clustered and calibrated low p_{T} jets as inputs. The p_{T} of the jets used as inputs are less than the large-R jet requirement of 200 GeV. The jet mass calibration can be an issue if a single jet with $R = 0.4$ is used as an input; with a fixed cone radius, there is no way around this problem. DR jet clustering is when the radius of the cone is determined by kinematic and angular features inside an evolving jet. Instead of just having a fixed radius, it has an extra term which uses a pT-weighted deviation of the distance between the elements that make up an intermediate jet.

The implementation of a variable-radius (VR) [101] jet algorithm was initially prioritised but later used to benchmark against alternative algorithms for comprehensive testing and validation. The inspiration was taken from a fat b-jet analysis using old and new clustering algorithms in new Higgs Boson searches at LHC [102]. In the study they compared jet clustering, VR and the fixed cone radius algorithm using Type II 2HDM in a fully hadronic final state. The search was previously done using slim b-jets; as mentioned, it is the anti-k$_T$ with a fixed cone radius of $R = 0.4$. When it was implemented, the VR algorithm had better signal sensitivity and reconstruction of Higgs mass peaks. Instead of having a fixed radius, the effective radius would have to be calculated using the equation:

$$R_{eff}(p_{Ti}) = \frac{\rho}{p_{\mathrm{T}}} \tag{3.6}$$

In simpler terms, how it works is that objects with higher p_{T} will be suppressed, and these objects will most likely be classified as jets. For lower p_{T}, these are combined with the nearest neighbour. Great technique to avoid events with very wide jets at low p_{T}. As seen in equation (3.6), the dimensionless input parameter, ρ, can be scanned to optimise the maximum desired sensitivity. To find the optimal value for ρ when using the VR algorithm, you can use the p_{T} distribution. The ρ should have the same order of magnitude as the events in the p_{T} distribution. One of the other advantages to the VR algorithm is that you can choose to use clustered jets or track jets as inputs, which has been done in [89], depending on which better complements the signature being explored. For the SVJ-b, it is worth noting that jets constructed from charged particles do not complement SVJs, so clustered jets were used as an input. So anti-k_t jets with $R = 0.4$ were used as input jets. The b-jet multiplicity for various ρ values was plotted, and it was found that the value that had clustered more b-jets was $< 2p_{\mathrm{T}}$. This value was deemed justifiable because the leading jet p_{T} was required to be 250 GeV or more.

In Figure 3.14, the histogram is depicted for $R_{inv} = 0.67$ and $R_{inv} = 0.33$, as labelled in the caption for the b-jet multiplicity before pre-selection requirements. Accessing the fixed cone radius algorithms, namely J04 and J10, which correspond to the anti-k_t algorithm for $R = 0.4$ and $R = 1.0$, it is observed that for most events there is a b-jet multiplicity of 0, following a step-like trend which is expected. The same applies for VRJ for $R_{inv} = 0.67$, but its multiplicities > 1; it reports higher events than J04 and J10. In the case of VRJ, for $R_{inv} = 0.33$, most events have a b-jet multiplicity of 2 . For DRJ, for both values of R_{inv}, most events have a b-jet multiplicity of > 1.

In Table 3.4, presented is the efficiency of signal selection for using the various jet clustering algorithms after the described pre-selection requirements. From the results, the conclusion that can be drawn is that VRJ and DRJ significantly enhance the signal efficiency when compared to J04 and J10. This is expected, as previously mentioned, that SVJs have a gap between the jets, so J04 would not be the ideal fixed radius to use. On the other hand, as mentioned again, J10 is used for boosted topologies, so there is a 200 GeV p_{T} requirement. Since a metric was created to test and compare the algorithm, once the SVJ pre-selection requirements are applied, which require a leading jet p_{T} of 250 GeV and scalar sum p_{T} of jets to be 600 GeV, it would decrease the values seen in Table 3.4.

To get a better visualisation of various objects and jet clustering, the $\eta - \phi$ was plotted as seen in Figure 3.15. The orange line represents the p_{T}^{miss} direction, the blue dots are stable b-hadrons, and

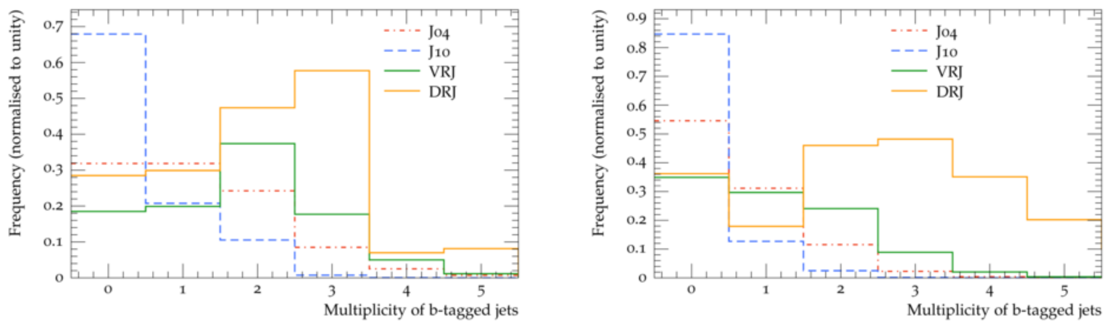

Figure 3.14: The distributions depict the b-jet multiplicity for jets clustered using anti-k_t fixed cone radius of 0.4 (J04) and radius of 1.0 (J10), variable-R jet algorithm (VRJ) and dynamic radius jet algorithm (DRJ). The first distribution on the left-hand side shows $R_{inv} = 0.33$ and the right-hand side $R_{inv} = 0.67$. The distributions are taken from [103].

	Selection Efficiency in %	
Selection	Signal $R_{\text{inv}} = 0.33$	Signal $R_{\text{inv}} = 0.67$
J04	33	12
J10	11	3
VRJ	60	35
DRJ	81	66

Table 3.4: Table showing the selection of jets clustered using different jet algorithms namely: anti-k_t fixed cone with $R = 0.4$ (J04) and $R = 1.0$ (J10), variable-R jets (VRJ) and dynamic-radius jets (DRJ). The selection efficiency is presented in percentage form for different R_{inv} fractions. Table is taken from [103].

the magenta dots are dark hadrons; the grey dot represents the widely used anti-k_t jets with a radius of 0.4, and the green circle represents VR jets. As depicted in Figure 3.15, at least one of the jets is in the direction or within a $\Delta\Phi$ of < 2 of the p_{T}^{miss} direction. The last diagram on the right-hand side supports the argument presented of using a different jet clustering algorithm, as seen there are two VR jets shown as two green circles but one grey dot. So in the pre-selection requirements, the event would have been discarded.

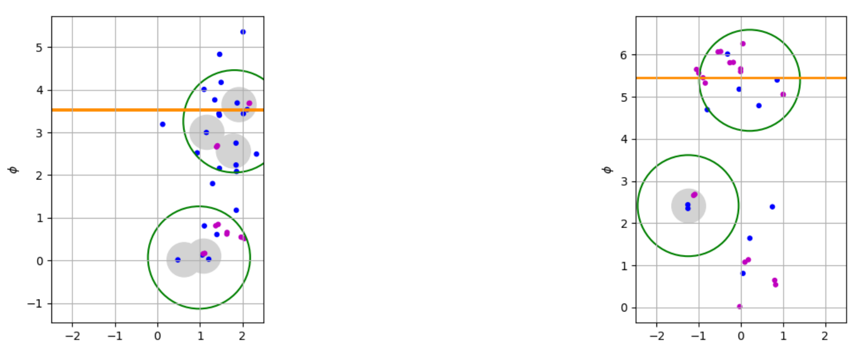

Figure 3.15: A diagram showing the $\eta - \phi$ plane with VR jets and anti-k_t. The orange line represents the p_{T}^{miss} direction, the blue dots are stable b-hadrons and the magenta dots are dark hadrons, the grey dot represents the widely used anti-k_t jets with radius of 0.4 and the green circle represent VR jets. Diagram can also be found in [103]

Looking at Table 3.4, the options for jet clustering are VR and DR, which both have shown remarkable improvement in selection efficiency. In Figure 3.14 and 3.4, examining the distributions closely and the numbers provided for DR jets shows that it would be the better option for improving signal retention. In the description of the VR jet clustering algorithm, it was mentioned that one of the advantages is that it can use track-jets or already calibrated jets as an input, depending on which complements the signature being explored. It was briefly mentioned that already calibrated jets encompass the SVJ signature better than using track jets. In the case of DR jet clustering, it does offer this option even though it is not supported, but at the cost of large systematic uncertainties, whilst using track-jets means that the differences in data and simulation are already accounted for. In Figure 3.16, further support for the conclusion that anti-k_t with $R = 0.4$ is not suited for SVJ is provided, as the effective radius was found to be $R = 0.6$ for DR jets. With all this said, it can be concluded that VR jet clustering is more sensitive to SVJ-b.

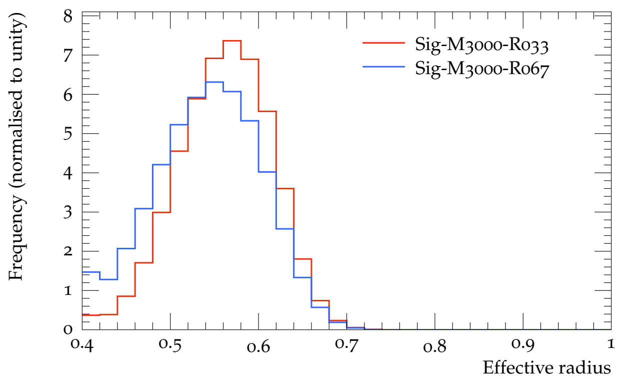

Figure 3.16: Distribution showing the effective radius for DR jet clustering algorithm using different R_{inv} values

3.7 Search Strategy

The remaining parts of the phenomenology study on the SVJ-b signature can be found in the paper [103]. Instead of doing a search for the signature, it was seen as beneficial to start off with a feasible study to get a sense of the topology and possible kinematic features. As noted in the above section, a set of well-thought-out signal reconstruction requirements were determined to better reconstruct the SVJ-b. Ultimately, the aim was to propose a search strategy with specific requirements, but this part would be challenging, especially with difficulty in modelling E_T^{miss} at the particle level. A more realistic approach would be to suggest discriminating variables and requirements to reduce the background. The discriminating variables that can be plotted are as follows:

1. H_T, which is defined as the scalar sum of p_{T} of the b-tagged VR jets in the event.
2. E_T^{miss}, is the magnitude of missing transverse momentum.
3. Effective mass, defined as a scalar sum of E_T^{miss} and H_T.
4. Lead jet p_{T}, this is the p_{T} of the leading b-tagged VR jet.
5. Number of sub-jets of radius 0.2 which can formed using anti-k_t algorithm from the leading b-tagged VR jet
6. $\Phi_{MET,closest-jet}$, defined as the $\Delta\Phi$ of the b-tagged VR jet closest to the direction of the missing transverse momentum.
7. p_T^{bal}, is the p_{T} balance between the closest jet and farthest b-tagged VR jets from p_{T}^{miss} in azimuthal direction, termed j_1 and j_2, defined as,

$$p_{\mathrm{T}}^{bal} = \frac{|\vec{p_{\mathrm{T}}}(j_1) + \vec{p_{\mathrm{T}}}(j_2)|}{|\vec{p_{\mathrm{T}}}(j_1)| + |\vec{p_{\mathrm{T}}}(j_2)|} \tag{3.7}$$

8. log (M_{bb}/p_T^{bb}), defined by logarithm of invariant mass of the two b-tagged VR jets over their invariant p_{T}

Looking at some of the kinematic distributions of these discriminating variables, in Figure 3.17, the plot for E_T^{miss} steeply declines, and for $E_T^{miss} > 800$ GeV, there is little to no background remaining. The proposed $E_T^{miss} > 600$ GeV requirement works well, as it reduces a significant amount of background without losing signal events. A similar observation can be made about H_T; some background steeply declines, like multijet $Z(\nu\nu)b\bar{t}$ for $H_T > 2500$ GeV, with little to no background remaining. The proposed $H_T > 600$ GeV requirement works well as it reduces a significant amount of background without losing signal events, similar to E_T^{miss}. The discriminating variable p_{T}^{bal} is one of the variables that are unique to the SVJ signature. There are no requirements implemented using this variable, but it assists in understanding the kinematics of the signature, and comments can be made on how balanced the signal or background is. In Figure 3.18, it is observed that multijet, $t\bar{t}$(had), including signals with $R_{inv} = 0.33$, are imbalanced compared to $R_{inv} = 0.67$ and remaining background processes.

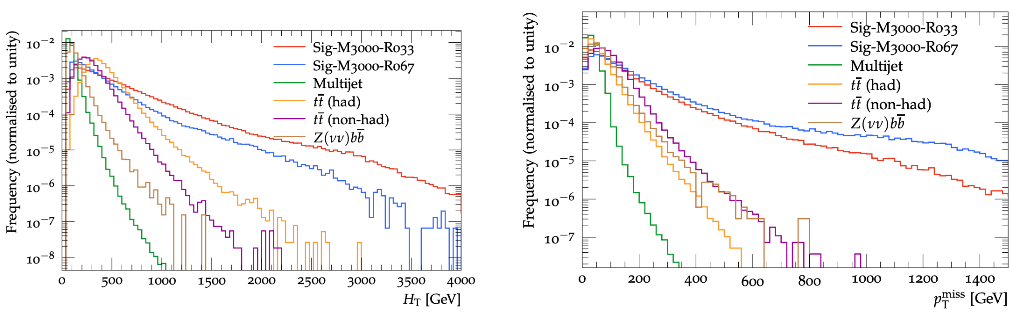

Figure 3.17: On the left-hand side is the kinematic distribution of the E_T^{miss} and on the right-hand side is the distribution for H_T. On both plots, all the dominating processes are present. The green is multijet, yellow is $t\bar{t}$ (had), purple is $t\bar{t}$ (non-had) and the gold is $Z(\nu\nu)(b\bar{b})$. The red and blue line are the signal events with different R_{inv} values

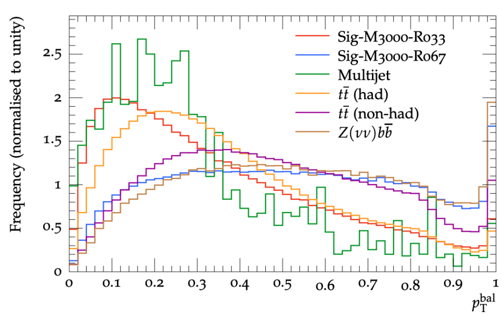

Figure 3.18: This is the kinematic distribution for p_{T}^{bal}. All the dominating processes are present in the plot. The green is multijet, yellow is $t\bar{t}$ (had), purple is $t\bar{t}$ (non-had) and the gold is $Z(\nu\nu)(b\bar{b})$. The red and blue line are the signal events with different R_{inv} values

The requirements that can be implemented to assist in reducing background are motivated by the ATLAS t-channel search for SVJ [74] and the discriminating variables that have been suggested. By requiring events to have H_T and $E_T^{miss} > 600$ GeV, at least 2 b-tagged jets that must lie within $\Delta\Phi < 2$ between the closest b-tagged jet and p_{T}^{miss} direction and implement a charged lepton veto. In Table 3.5, presented are the yields after each selection was implemented, which is intended to give an idea of how effective each cut is in reducing the background and how it affects signal acceptance. The lepton veto and requirement of at least 2 b-tagged VR jets and the closest b-tagged jet to be within $\Delta\Phi < 2$ of the direction of the missing transverse momentum reduced most of the background by 10^1 or 10^2 at most. The $E_T^{miss} > 600$ GeV seemed to reduce the background yields to 0 except for $t\bar{t}$ (non-had), which still had yields with two orders of magnitude. This also proportionally affected the

	Yields after each selection step					
Selection	Signal R_{inv}= 0.33	Signal R_{inv}= 0.67	Multijet	$t\bar{t}$ (had)	$t\bar{t}$ (non-had)	$Z(\nu\nu)bb$
Start	112439	112439	3×10^{16}	3.9×10^{6}	1.2×10^{5}	1.4×10^{5}
Lepton Veto	110521	111506	3×10^{15}	3.7×10^{6}	1×10^{5}	1.4×10^{5}
Two b-tagged VR jets	65615	38234	1.5×10^{14}	3.4×10^{6}	9.4×10^{4}	22640
$\phi_{\text{MET,closest-jet}} < 2$	38824	15783	6×10^{13}	2.4×10^{6}	4.5×10^{4}	2626
$E_{\text{T}}^{\text{miss}} > 600$ GeV	1604	1316	0	0	292	0.1
$H_{\text{T}} > 600$ GeV	1246	673	0	0	208	0.1

Table 3.5: A table showing yields for two SVJ-b signals (corresponding to a mediator mass of 3000 GeV and R_{inv} of 0.33 and 0.67) and all leading background processes for a representative set of requirements. The numbers are normalised to 139 fb^{-1} of integrated luminosity.

signal yields for $R_{inv} = 0.33$ and $R_{inv} = 0.67$, with $R_{inv} = 0.33$ having three orders of magnitude worth of events after the last cut and $R_{inv} = 0.67$ having two orders of magnitude.

Chapter 4

Measurement of b-jet cross-section in ATLAS

In particle collisions, b-quarks are produced in abundance, and their hadronization leads to the formation of b-jets, which are key signatures in many physics analyses. When it comes to b-jets, they are incredibly useful in many final states, for example, in phenomenology studies exploring new BSM signatures such as SVJ-b, performance studies which will be covered in Chapter 5, and measurements which will be presented in this Chapter. Additionally, b-jets play an essential role in measurements of SM processes, such as $H \rightarrow bb$ and top quark decays, where the identification of b-jets helps determine the underlying kinematics and validate theoretical predictions.

4.1 Measurements Validating the Standard Model

One of the major successes in particle physics is the SM. In the introduction, a thorough discussion on the different fundamental particles in the SM and how they interact with one another and nature was presented. It was also described as a theoretical foundation constructed from observations that speculates and fits new data and a comprehensive framework for understanding the basic building blocks of the universe and their interactions. Given its huge success for being in agreement with data, it is reasonable to assume the SM might be missing some particles that could explain important phenomena such as DM. This will bring us to the motivation behind including E_T^{miss} in the measurement of the b-jet cross-section.

The ATLAS Detector at the LHC has played a pivotal role in testing and validating the SM at new centre of mass energies. As mentioned, the detector was built with a goal to study the fundamental particles and forces described by the SM and to search for BSM physics. So far, ATLAS has achieved significant milestones, particularly through its discovery of the Higgs boson and by providing precise measurements of known particles and interactions. At the same time, ATLAS is constantly searching for new physics, aiming to uncover phenomena that the SM cannot explain. Every run, the detector is upgraded to reach a higher centre of mass and integrated luminosity to increase the chances of observing new physics. This means that new measurements of particle properties need to be done to help confirm that if at higher energies the SM remains valid and new searches to see if at higher energies ATLAS has excluded many potential models which may have set constraints on the models.

Using the data collected from the detector, either a measurement of particle properties can be done or a search for new phenomena which would be part of BSM physics. When doing a measurement, the aim is to test the SM predictions by precisely measuring the properties of the particles and interaction that are already theoretically understood and proven by the SM. In a measurement, a specific final state that has been observed is measured, in which a comparison is done between the data and Standard Model predictions given by Monte Carlo event generators. There are few properties that can be precisely measured, such as cross-sections, particle decay rates, masses, coupling constants, and other kinematic observables. Doing measurements every run, after the ATLAS detector has undergone upgrades, helps to refine the understanding of the Standard Model and test its validity at new integrated luminosity and energies. This is just to emphasise that measurements are essential for improving the accuracy of theoretical models and confirming the consistency of the Standard Model with experimental data.

A thorough literature review yielded a recent measurement of an inclusive differential cross-section with missing transverse momentum and jets [104], but there is no measurement of the b-jet cross-section using data collected by the ATLAS detector in Run 2 of the Large Hadron Collider. The results presented are a measurement of the b-jet cross-section as a function of E_T^{miss} using data collected by the ATLAS detector at a centre of mass of 13 TeV using the full 140 fb^{-1} of integrated luminosity in Run 2 of the LHC. A high E_T^{miss} region was defined, and various kinematic variables are plotted in this region, comparing data against SM predictions given by Monte Carlo event generators.

In Section 2 and 3, E_T^{miss} has been defined and discussed. The origin of E_T^{miss} is linked to neutrinos, which can travel through the various layers of the detector without interacting with any calorimeter. As a result, they are undetected. It has become an important collider signature for BSM physics, as it alludes to the presence of dark matter. The inclusion of missing transverse momentum in measurements can make them sensitive to dark matter and other beyond the Standard Model signatures. In [104], they demonstrate that measurements can be used in searches and setting constraints without having to do extra work in computing detector simulations. Typically, when taking the ratio of the cross-section of the observable you are measuring, it should cancel out [104], but in this case, if there were to be a DM contribution in E_T^{miss}, as expected, it would not cancel out. Since DM is part of BSM signatures and would not be accounted for or validated by SM theoretical predictions. Ultimately, measurements done with discriminating variables that are sensitive to BSM searches can be used in recasting, which helps to learn about different models.

In this case, the inclusion of E_T^{miss} in this measurement is motivated by this. The main purpose of the measurement is to confirm the consistency of the SM with experimental data collected by the ATLAS detector. Another objective would be to explore and learn more about the SVJ-b signature. The measurement is done in a high E_T^{miss} region; to be precise, events to have $E_T^{miss} > 300$ GeV were required, amongst other event selections. The reason for this is motivated by SVJ-b, knowing that at detector level, E_T^{miss} is not only attributed to the detection of neutrinos but also mismeasurements and detector effects that also contribute to these observables, so requiring a higher E_T^{miss} assists in removing their contribution to E_T^{miss}. If there is still a significant contribution of E_T^{miss} after accounting for neutrinos and mismeasurements, it could indicate the presence of DM. Some of the measured observables that are presented are taken from the suggested discriminating variables in SVJ-b. From the measurement, it can be observed what contribution is from the SM, so when looking at the results from SVJ-b, there is an idea of what contribution is from SM and possibly BSM. Overall, including discriminating variables that are linked to BSM signatures such as E_T^{miss} can make measurements sensitive to dark matter and BSM searches.

4.2 Theoretical Predictions and Monte Carlo Simulations

For the unfolding, data and simulated Monte Carlo (MC) event samples at both particle- and detector-level are used. The data collected during the hard scattering of protons, known as *pp* collision. The protons are collided in bunches every 25 ns. The data used in the unfolding was collected by the ATLAS detector at full luminosity of 139 fb^{-1} with a centre of mass $\sqrt{s} = 13$ TeV during Run 2, which was in the time period 2015-2018.

The dominant standard model contributions were identified to be multijet and $t\bar{t}$. The $t\bar{t}$ contributions were divided into all-hadronic (all-had) and non-all-hadronic (non-had) decay modes. The decay mode for all-had is $q\bar{q}b$, which has the highest branching ratio. For non-had, there are two decay modes, namely, semi-leptonic decay (lvb, $q\bar{q}b$) and dileptonic decay (lvb, lvb). The dileptonic decay products can be classified into lepton of the same flavour or mixed. For the multijet and $t\bar{t}$ MC event samples, the hadronization, parton showering, and underlying events were all simulated using `PYTHIA` 8.235 [12], which generates events using the leading-order (LO) $2 \to 2$ matrix element. For the parton distribution function, NNPDF2.3LO [105] was used with a set of generator parameters named A14 tune. The decay of heavy flavours, namely c- and b-hadrons, was handled by `EVTGEN` 1.6.0 [106]. For the $t\bar{t}$ event samples, the parton showering was interfaced with `POWHEG-BOX` v2. A second set of MC event samples was used for the unfolding. This is beneficial in visualising and understanding the impact of modelling uncertainties in event generators. The two sets of samples would have different underlying approaches to showering processes. The simulated events for the second sample for $t\bar{t}$ were generated using `SHERPA` 2.2.12 [107]. Compared to `PYTHIA`, which generates events using $2 \to 2$ matrix elements for processes, `SHERPA` uses multi-leg $2 \to N$ matrix elements. The parton distribution function used was NNPDF3.0NNLO [105] with the generator parameters set to `SHERPA`.

For the MC simulated particles at detector level, the event samples are passed through `GEANT4`, which models the ATLAS detector [108] and simulates detector effects.When the protons are hard scattered close to the speed of light, there are particles that collide with each other within the bunches. When the detector measures multiple proton-proton collisions in a single bunch crossing, it is called pile-up. The pile-up is proportional to the high luminosity and cross-section. In this measurement, the pile-up was modelled using using `PYTHIA` with the A3 tuning parameter and overlaid with the hard scattering events.

4.3 Event Selection and Object Reconstruction

The details pertaining to how the *b*-jets and E_T^{miss} were reconstructed for this measurement can be found in Chapter 2.3. The reconstructed jets also undergo calibration; this process is merely implementing corrections to the jets to account for differences between data collected by the ATLAS detector and MC event generators [109]. This ensures that the jets used in the measurement are correctly described. With hadrons being collided at high speed and large integrated luminosity, the consequence for this is there will be a large amount of pile-up. Some of the jets can originate from a pile-up. It is easier to identify when electrons or muons are mistaken as jets since they are associated with a collision using tracks. Whilst hadrons can originate from multiple interactions. To ensure that jets from pile-up are not included, jet vertex tagging is implemented. The tracks associated with jets are used to calculate the fraction of energy from the hard scattering or pile-up interactions [110].

Events were selected using the un-prescaled E_T^{miss} trigger with the lowest threshold [111]. The trigger used was taken from [74]. In the year 2015, the threshold was set to 70 GeV, and in the years 2016-2018, it was increased to reach 110 GeV. The trigger used is only based on information and measurements from the calorimeter cluster cells and does not take into consideration the reconstruction of muons. Both detector- and particle-level jets were required to have $p_{\mathrm{T}} > 30$ GeV and be within $|\eta| < 2$. A veto on charged leptons with a p_{T} threshold of 7 GeV was implemented.

The goal of the measurement was to measure the cross-section as a function of high E_T^{miss} and other observables. The results presented are for events that passed the following requirements: events were required to have at least four jets; out of the four jets, two were required to be b-tagged jets. A leading jet $p_{\mathrm{T}} > 100$ GeV and $|\phi_{all,bjets} - \phi_{MET}| > 0.2$ selection was implemented. Even though the cut implemented on $|\phi_{all,bjets} - \phi_{MET}|$ does cut out the most sensitive region for SVJ-b, this is to avoid the region with the worst mismodelling and keep some data blinded if a search were to be done whilst understanding the background modelling in the most amount of phase space. Lastly, events used in the measurement were required to have $E_T^{miss} > 300$ GeV.

4.4 Measured Observables

When deciding which measured observables would be in the measurement, inspiration was taken from SVJ-b phenomenology study presented in Chapter 3. The reason for this ties with the explanation behind doing the measurement in a high E_T^{miss} region instead of it being inclusive. The inclusion of missing transverse momentum in measurements can make them sensitive to dark matter and other BSM signatures, such as SVJ-b.

The cross-section is a function of the following observables in the defined high E_T^{miss} region mentioned:

- E_T^{miss}, which is the magnitude of the missing transverse momentum
- N_{bjets}, the number of b-tagged jets constructed using anti-k_t algorithm with $R = 0.4$.
- N_{jets}, the number of jets constructed using anti-k_t algorithm with $R = 0.4$.
- Leading jet p_{T}
- H_T, the scalar sum p_{T} of b-tagged jets
- $H_{T,b}$, the scalar sum p_{T} of all jets
- $|\phi_{closest,bjet} - \phi_{MET}|$ which is the azimuthal angle between the closest b-tagged jet and missing transverse momentum.
- $|\phi_{all,bjets} - \phi_{MET}|$ which is the azimuthal angle between all b-tagged jets and missing transverse momentum.
- $\Delta\phi_{bjets}$ which can also be defined as $|\phi_{bjets,1} - \phi_{bjet,2}|$ is the azimuthal angle between all b-tagged jets
- $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$, is the azimuthal angle between leading b-jet and sub-leading b-jet
- $\Delta R_{b,b}$ which is the the angular distance between the b-tagged jets

The the angular distance between the b-tagged jets is calculated as follows:

$$\Delta R_{(b,b)} = \sqrt{\Delta\phi^2_{(b,b)} + \Delta\eta^2_{(b,b)}} \tag{4.1}$$

4.5 Unfolding

Unfolding is a term used when kinematic distributions are corrected for detector effects. The distributions are taken from detector-level to particle-level so that they can be directly compared to theoretical predictions. Particle-level refers to distributions obtained from events generated from MC event simulations such as Pythia. Whilst on the other hand, detector-level refers to when the generated events from MC event simulations are passed through a detector simulation, such as `GEANT4`, which mimics the behaviour of the detector as closely as possible. As expected, the results obtained from MC event generators will not perfectly match the results obtained from the detector, and unfolding is the procedure done to correct for this. It is done to account for limited detector acceptance, bin migration, inefficiencies of the detector, limited resolution, and any particles detected that are unaccounted for [7].

To explain unfolding in a mathematical and technical approach, to do a basic unfolding at least three distributions which have events that have passed certain selections and should measure a specific variable. The first distribution should contain detector-level events, also known as truth, and in the explanation, it will also be referred to as true value. As mentioned, detector-level events are generated using MC event generators and passed through a simulation that mimics the detector. For the detector-level distribution, which can also be referred to as truth distribution, let's say variable X is measured and plotted in a histogram with M number of bins. The variable, $\bar{\mu}$, represents the expected value of the number of entries in bin i of the distribution. Which can be mathematically expressed as $\mu_{total} = \sum_i \mu_i$ and $P(y|i) = P_i = \mu_i \ \mu_{total}$ [112].

The goal would be to not guess the parameters θ that parametrise $f(y|\theta)$ but to find the value of μ_i for each bin such that $\bar{\mu} = (\mu_i ... \mu_M)$ [112], where $\bar{\mu}$ is the expected number of events in M bins in the distribution if variable X was measured for each event. It is worth keeping in mind that this is not easy; the act of measuring any variable, including X, can include fluctuations which have to be considered and later accounted for. The second distribution needed for unfolding should contain particle-level events, also known as reconstructed events, and in the explanation will also be referred to as the measured value. As mentioned, particle-level events are generated using MC event generators. In the case of reconstructed distribution, the same variable X is measured with N number of bins, and the expected number of events in N bins is given as $\bar{n} = (n_i ... n_N)$ [112].

When variable X is compared, it is expected that the reconstructed and true values of this variable will be different by a certain value, and this value accounts for detector effects. The response matrix is constructed using the truth histogram and the reconstructed histogram and is denoted as R_{ij}. The definition of this matrix is: R_{ij} = P(reconstructed value in bin i— true value in bin j) [112], the index i will run from 1→ N and the index j will run from 1→ M. In statistical terms, the response matrix is the conditional probability of an event that is reconstructed in bin i of the data given that its true value was in bin j of the truth histogram [112]. The response matrix is an $N \times M$ matrix. It is important because it contains information that corrects for detector acceptance, efficiencies and bin migrations. Bin migration is when the same event ends up in different bins in particle- and detector-level distributions due to inaccuracy in the measurement. The relationship between detector-level distribution, particle-level distribution and response matrix can be mathematically expressed as:

$$f_{measured}(x) = \int R(x|y) f_{true}(y) dy \tag{4.2}$$

The third distribution needed for unfolding should contain events measured by the detector also known as data events and in the explanation will also be referred to as the expected value. The expectation values for data events are defined as $\nu = E(n) = (\nu ... \nu_N)$ [112] where n is the data collected using the detector and ν the expected data distribution. With all the needed distribution for unfolding defined, the unfolding problem can be mathematical expressed as:

$$\mu = R^{-1} \nu \tag{4.3}$$

The inverse of the response matrix can also be referred to as the smearing matrix; when multiplied by the data distribution, it corrects for detector effects, and the resulting histogram should be the truth distribution, in theory. This is often seen as an easy task because it just requires inverting the response matrix, but it is known to have a large variance and statistical fluctuation, which will be discussed in the following section.

There are a few techniques that can be implemented to do the unfolding; the results presented were obtained using Iterative Bayesian (IB) Unfolding implemented in the RooUnfold Framework [113]. In the RooUnfold Framework the unfolding was done using the function `RooUnfoldBayes`. The framework created the response matrices for each of the measured observables using the function `RooUnfoldResponse`, which requires the particle-level and detector-level distributions as inputs to have the same number of bins. The response matrix was used to correct for detector effects for events that passed both particle- and detector-level selections.

This technique was introduced by D'Agostini [114] and inspired by Bayesian statistics. This is an iterative method that calculates the maximum likelihood estimate of the matrix inversion. The response matrix has to be computed for each distribution in this technique. As mentioned, the response matrix gives the conditional probability for an event to be reconstructed in bin i given that its true value is in bin j. In Bayesian statistics it is presented as:

$$A_{ij} = P(G_j|W_i) \tag{4.4}$$

For these measurements, the response matrices were constructed for all the measured observables for events generated with the `PYTHIA` event generator and the `SHERPA` MC event generator. In Figures 4.1, 4.2, and 4.3 are the response matrices constructed for using `PYTHIA` particle- and detector-level events, and Figures 4.4, 4.5, and 4.6 for using `SHERPA` particle- and detector-level events using the `RooUnfoldResponse` function. The more diagonal these matrices are correlates to less bin migration observed. In the response matrices constructed for this measurement, even though there are diagonal entries, there is a lot of bin migration in the MC-generated events.

The Figures 4.3 and 4.6 correspond to the angular observables and, as shown, take up the entire accepted angular phase space. This means that the b-jets can be found in all directions, given that the results show bin migration. They are not likely to be in one particular direction in the detector. The Figures 4.1, 4.2, and Figures 4.4, 4.5 correspond to jet multiplicities and p_{T}-related observables which, when compared to the angular response matrices, only take up the lower diagonal of the matrix. Given that there is some bin migration, there is a specific p_{T} range that the b-jets have when looking at these observables. The bin migration observed in the p_{T}-related observables is less than the migration observed in angular observables because some of the p_{T}-related observables are closer to being diagonal than the angular observables.

Once the response matrix is created using particle-level distributions and detector-level distributions, the unfolding matrix is computed and denoted as C_{ij} in the equation below. The variable $P(W_i)$ [7] is the input prior and this technique is iterative and i denoted the number of data events in the particle-level distribution bin i.

$$C_{ij} = P(W_j|G_i) = \frac{P(G_j|W_i) \cdot P(W_i)}{\sum_i P(G_j|W_i) \cdot P(W_i)} = \frac{B_{ji} \cdot P(W_i)}{\sum_i B_{ji} \cdot P(W_i)} \tag{4.5}$$

D'_i is computed using the equation (13). D_j is the number of events measured in bin j. Using the response matrix created the unfolding matrix can be constructed for that distribution.

$$D'_i = \sum C_{ij} D_j \tag{4.6}$$

Using Bayes' Theorem the unfolding can be defined as the following:

$$C_{ij} = P(T_i|D_j) = \frac{P(D_j|T_i) \cdot P(T_i)}{\sum_i P(D_j|T_i) \cdot P(T_i)} = \frac{b_{ji} \cdot P(T_i)}{\sum_i b_{ji} \cdot P(T_i)} \tag{4.7}$$

The variable $P(T_i)$ is the input from the previous iteration. The technique uses the input from the previous iteration to limit the bias created by the trend in the distribution predicted by the simulation. The estimators from the previous iteration use a goodness of fit test; in this case, the χ^2 test is used in each iteration. Information on χ^2 can be found in [114]. The difference between the results for different iterations indicates the stability of the iterative procedure. Based on the number of iterations that was set, the difference between the number of events in data histogram bin i and the number of events in truth histogram bin j should be minimised. For this technique, it is worth mentioning that the default number of iterations for this method is 4. The uncertainties that come with using IB unfolding are discussed in the following section.

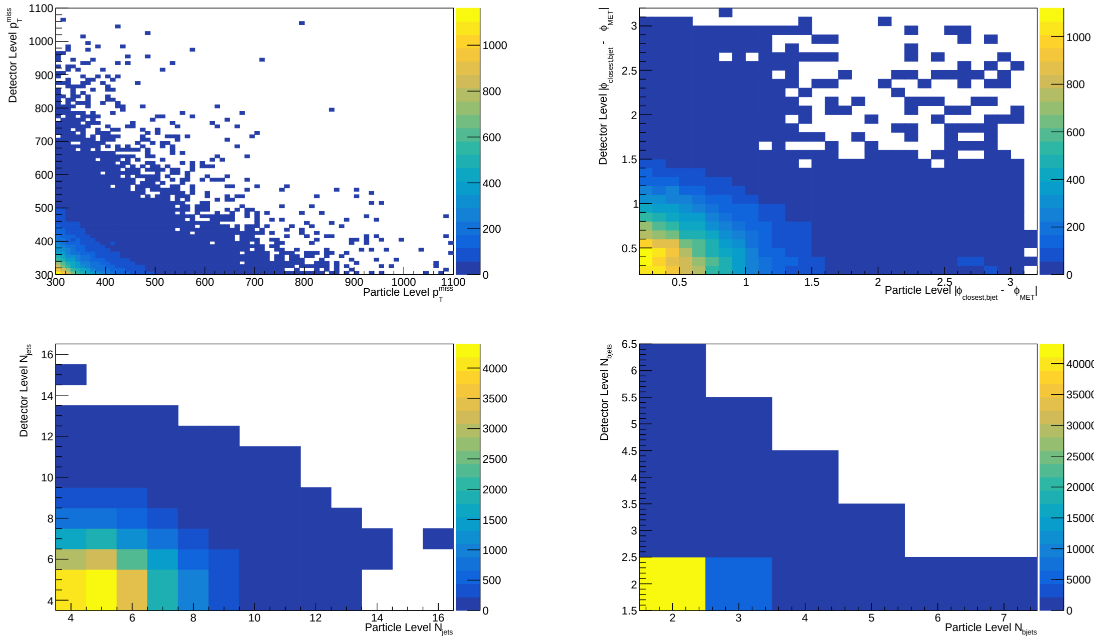

Figure 4.1: The response matrix is a $N \times M$ matrix. It contains information that corrects for detector acceptance, efficiencies and bin migrations. The following matrices were constructed using events generated using PYTHIA. In the top panel, on the left-hand side is the response matrix for p^T_{miss} and on the right-hand side for $|\phi_{closest,bjet} - \phi_{MET}|$. In the bottom panel, on the left-hand side is the response matrix for N_{bjet} and on the right-hand side N_{jets}.

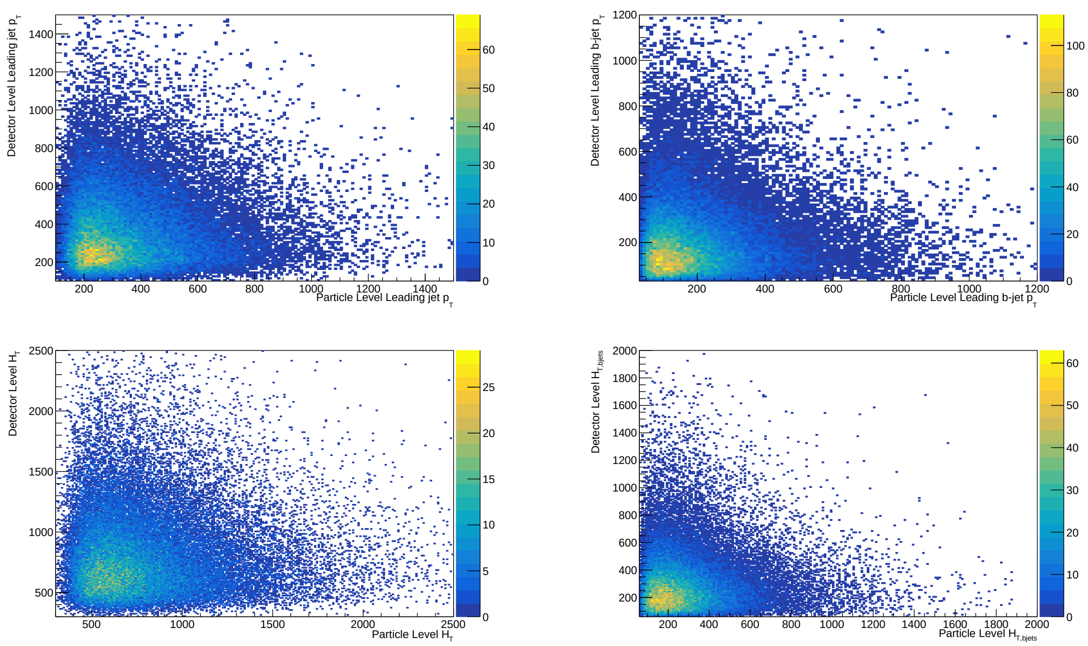

Figure 4.2: This response matrix contains information that corrects for detector acceptance, efficiencies and bin migrations. The following matrices were constructed using events generated using PYTHIA. In the top panel, on the left-hand side is the response matrix for leading jet p_{T} and on the right-hand side for leading jet p_{T}. In the bottom panel, on the left-hand side is the response matrix for H_T and on the right-hand side $H_{T,bjets}$.

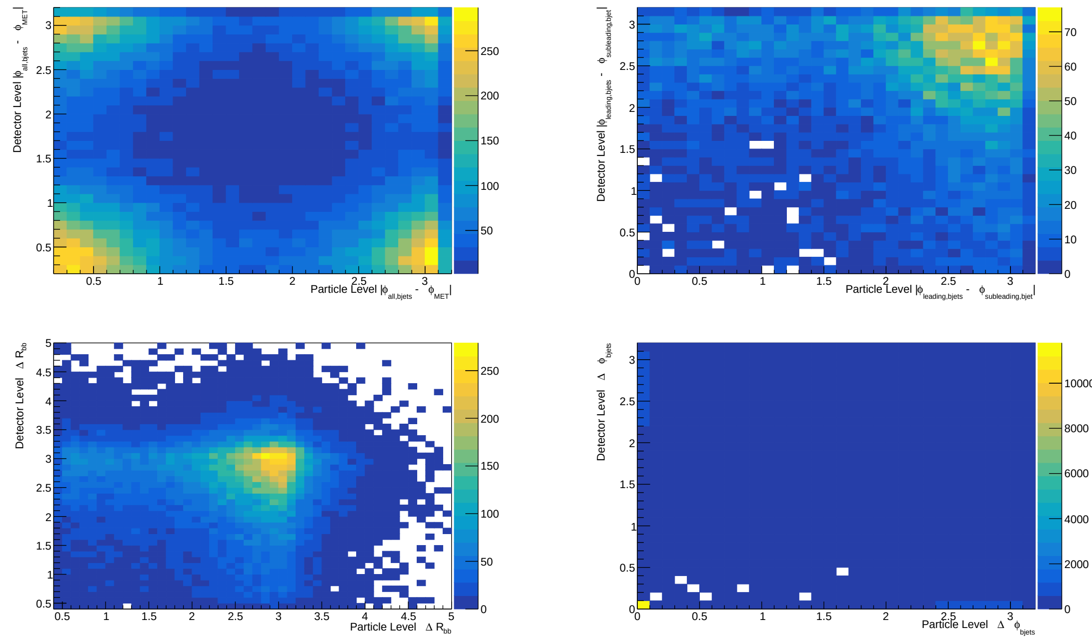

Figure 4.3: This response matrix contains information that corrects for detector acceptance, efficiencies and bin migrations. The following matrices were constructed using events generated using PYTHIA. In the top panel, on the left-hand side is the response matrix for $|\phi_{all,bjets} - \phi_{MET}|$ and on the right-hand side for $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$. In the bottom panel, on the left-hand side is the response matrix for $\Delta R_{b,b}$ and on the right-hand side $\Delta\phi_{bjets}$.

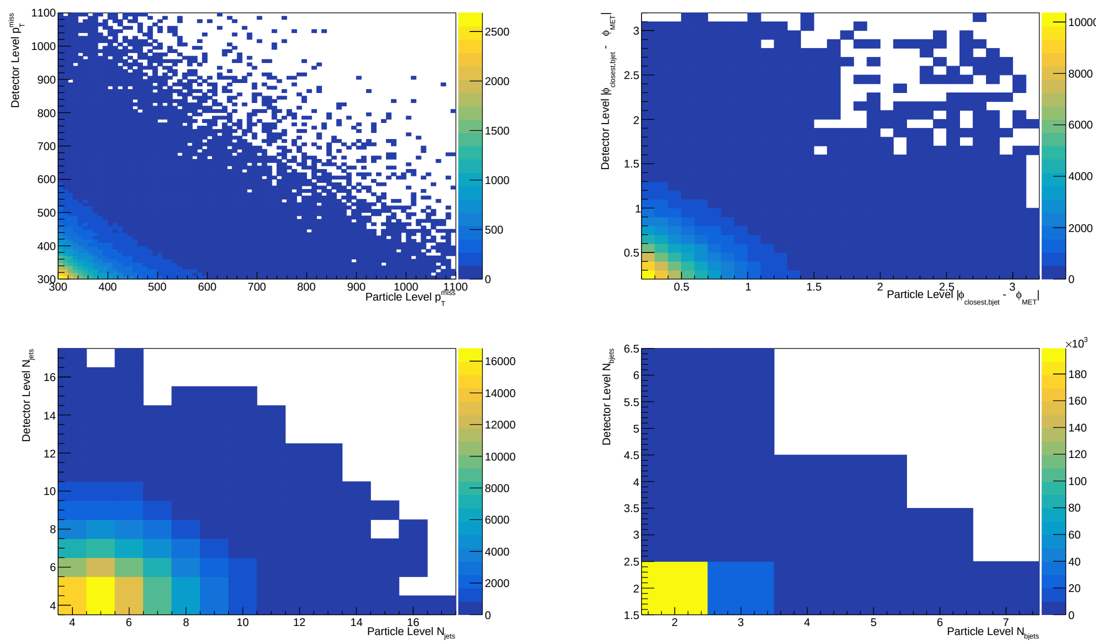

Figure 4.4: This response matrix contains information that corrects for detector acceptance, efficiencies and bin migrations. The following matrices were constructed using events generated using SHERPA. In the top panel, on the left-hand side is the response matrix for p^{T}_{miss} and on the right-hand side for $|\phi_{closest,bjet} - \phi_{MET}|$. In the bottom panel, on the left-hand side is the response matrix for N_{bjet} and on the right-hand side N_{jets}.

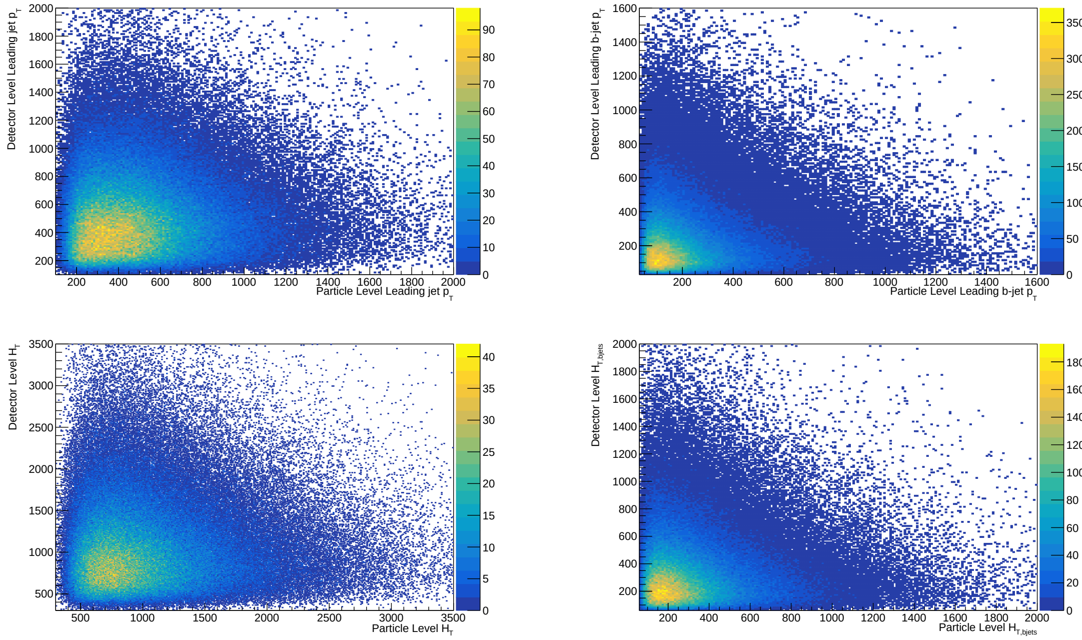

Figure 4.5: This response matrix contains information that corrects for detector acceptance, efficiencies and bin migrations. The following matrices were constructed using events generated using SHERPA. In the top panel, on the left-hand side is the response matrix for leading jet p_{T} and on the right-hand side for leading b-jet for p_{T}. In the bottom panel, on the left-hand side is the response matrix for H_T and on the right-hand side $H_{T,bjets}$.

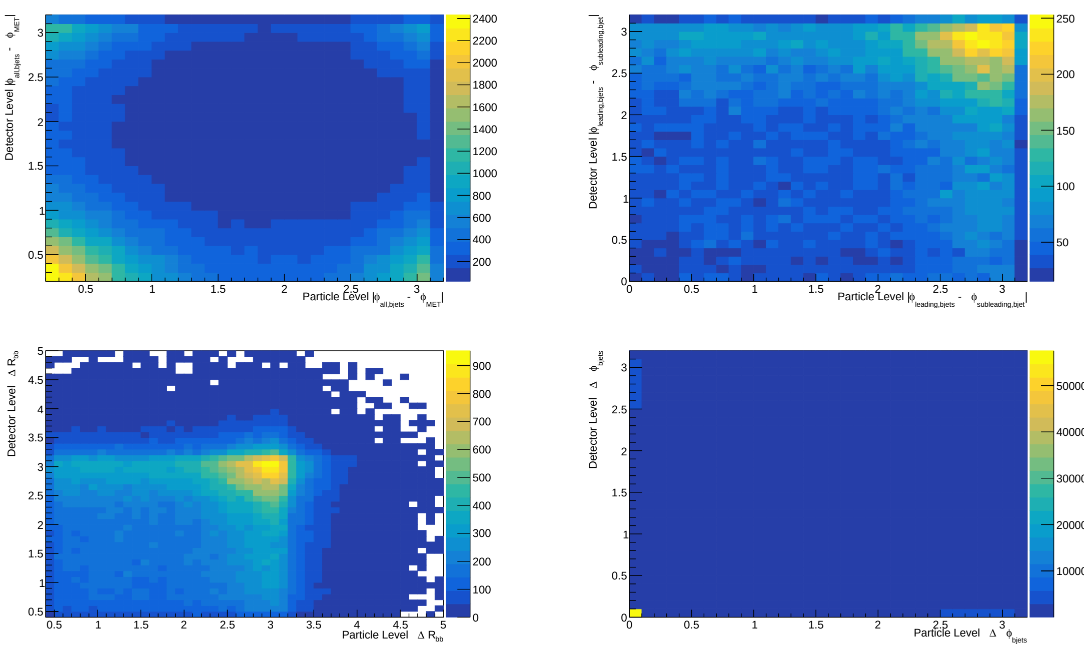

Figure 4.6: This response matrix contains information that corrects for detector acceptance, efficiencies and bin migrations. The following matrices were constructed using events generated using SHERPA. In the top panel, on the left-hand side is the response matrix for $|\phi_{all,bjet} - \phi_{MET}|$ and on the right-hand side for $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$. In the bottom panel, on the left-hand side is the response matrix for $\Delta R_{b,b}$ and on the right-hand side $\Delta\phi_{bjets}$.

4.6 Systematic Uncertainties

4.6.1 Detector and Experimental Uncertainties

No experimental results are infinitely accurate or precise. All experiments come with experimental and statistical uncertainties. Every object reconstructed from the detector has an uncertainty associated with it, and this uncertainty gets propagated into the measurement during the unfolding.

When reconstructing jets, the energy deposits in the calorimeter cell clusters are measured and used to construct jets using a jet clustering algorithm. There is an uncertainty associated with the jet energy scale and jet energy resolution [115] due to limited detector resolution. The uncertainty associated with jet energy scale can be determined by comparing jets constructed using calorimeter cluster cells and jets constructed from tracks in the detector from inclusive dijet events.

The track reconstruction efficiency, fake tracks and track parameter scales and resolution are uncertainties associated with tracking and identification of final state particles in the detector. The main contribution to inefficiency in tracking is the multiple scattering in the inner detector; the uncertainty arises depending on whether the correct inner detector material is simulated [116]. Another contribution to the inefficiency is certain places in the detector; there tend to be high multiplicity environments due to the silicon pixel and micro-strip cluster merging, so the modelling in these high-density regions can be poor and affect the accuracy in track reconstruction [117]. When there are multiple charged particles in a final state, some tracks tend to be falsely reconstructed and identified. This can be easily reduced by ensuring that the track selection is well-suited for the final state particle that you want to measure. Lastly, since the measurement is for a b-jet, the track parameters that are important are p_{T} and d_o [118, 119]. The uncertainty in reconstructing and measuring these parameters is propagated into results obtained in the measurement.

When it comes to b-tagging performance, the b-jets as mentioned are calibrated and classified as b-jets at 77% efficiency, which means the b-tagging performance has some uncertainties which can be found in [120]. To correct for flavour tagging, a scale factor is computed using b-jets in data and MC samples. The uncertainties in b-jet efficiency and c-jet misidentification are determined by using $t\bar{t}$ samples [121], and for the uncertainty in light jets, the misidentification is determined using dijet events. The uncertainty for track jets with $p_{\mathrm{T}} > 300$ GeV is extrapolated using low p_{T} events since there are not enough statistics to get an accurate value for the associated uncertainty. This means that the uncertainty in the impact parameter is propagated into the extrapolation for high p_{T} events.

Other uncertainties that need to be taken into account are trigger inefficiencies, which can be found [122]; uncertainties in the integrated luminosity for Run 2 data, which is determined using the LUCID-2 detector [123, 46]; and the reconstruction of E_T^{miss} [54].

4.6.2 Unfolding and Theoretical Uncertainties

The uncertainty in unfolding is introduced by the technique used and the inputs. Since it uses reconstructed objects from the detector, some of the uncertainties discussed above are expected to be carried into the unfolding. The theoretical uncertainties are associated with MC samples. The response matrix is one input that contributes to the unfolding uncertainties, even though it contains information that corrects for detector acceptance, efficiencies and bin migrations. In the results presented, the uncertainties mentioned in 4.6.1 are not explicitly accounted for; only the dominant modelling uncertainty is accounted for.

The theoretical uncertainties as mentioned come from the MC samples used, specifically from the choice of PDF set parameters and tune, renormalisation, factorisation scale and reweighting procedure. These uncertainties are grouped together and called systematic uncertainties. Introducing a different MC event generator such as `PYTHIA` and `SHERPA`, the difference in the unfolding results of the two MC event generators is treated as systematic uncertainties. In theory, the unfolding results of the two are meant to be the same.

The unfolding uncertainties associated with the technique implemented in RooUnfold are discussed in [113]. Using IB unfolding is commonly used in ATLAS but can come with a large statistical fluctuation due to incomplete regularisation. What regularisation does is it forces the fluctuations to smoothen out but can introduce a large bias. The bias is defined to be the difference between the estimator's expectation and the true value. Even though the iterations are a form of regularisation, it is incomplete because IB unfolding interrupts the maximum likelihood before it reaches its true maximum. Techniques similar to IB unfolding with regularisation are called Single Value Decomposition (SVD), which is also implemented in RooUnfold, the difference being it focuses on suppressing statistical fluctuations.

A parameter that also has an associated uncertainty is i, the number of iterations. In the RooUnfold framework, as mentioned, the recommended or default number of iterations is $i = 4$. To minimise the uncertainty introduced, a balance would need to be found, as the higher the number of iterations, the more statistical uncertainty is introduced in the unfolding; on the other hand, fewer iterations lead to a stronger bias from the input prior.

4.7 Results

Prior to the unfolding, the SM processes that are known to result in a b-jet final state were plotted using `PYTHIA` MC samples at reco level alongside uncorrected data collected from the detector. The $t\bar{t}$ sample was divided into non-all hadronic and all hadronic. The distributions shown in Figures 4.7, 4.8, 4.9, 4.10, and 4.11 were initially plotted to confirm the leading SM processes contribution in a b-jet final state. The results confirmed two things. It confirmed that $t\bar{t}$ (non-had) had a leading SM contribution for a b-jet final state and that no SM process was excluded or unaccounted for.

Figure 4.7 shows the E_T^{miss}. For this observable, the events were required to at least have $E_T^{miss} >$ 300 GeV; as mentioned, the selection was motivated by SVJ-b. The distribution had an expected trend. The b-jet and jet multiplicity were plotted in Figure 4.8. Variables associated with p_{T} were plotted and are shown in Figure 4.9, and the angular variables were plotted and shown in Figure 4.10. The variables linked to p_{T} were plotted for all jets and b-jets, and it is observed that they peaked at different values. This is expected, as all jets included b-jets, so it is likely that the distributions will peak at higher p_{T} for all jets. This can be observed in the leading jet p_{T} and leading b-jet p_{T} distribution; the same applies for H_T and $H_{T,bjets}$. There is also a noticeable increase in statistical fluctuations as the distribution steeply declined in Figure 4.9.

For the angular functions, the observables that would display spatial properties of b-jets in relation to each other and E_T^{miss} were plotted. The inclusion of E_T^{miss} in $\Delta\phi$ variables is to make it easier to understand the measurement in relation to SVJ-b. In Figure 4.10, it is observed in $\Delta R_{b,b}$ that most b-jets would be back-to-back, and the smallest angular distance they can have is $R = 0.4$ due to the chosen radius in the anti-k_t algorithm. For the $|\phi_{all,bjets} - \phi_{MET}|$ distribution in Figure 4.10, it is understood that b-jets in relation to E_T^{miss} direction are either back-to-back or within $\Delta\phi < 0.5$ of each other, and most leading and sub-leading b-jets are just back-to-back, as shown by $|\phi_{leading,bjet} - \phi_{subleading}|$ in Figure 4.10. Lastly, in Figure 4.11, the variable plotted was also taken from the recommended discriminating variables for SVJ-b. The closest b-jets in relation to the direction of E_T^{miss} lie within $\Delta\phi < 1$.

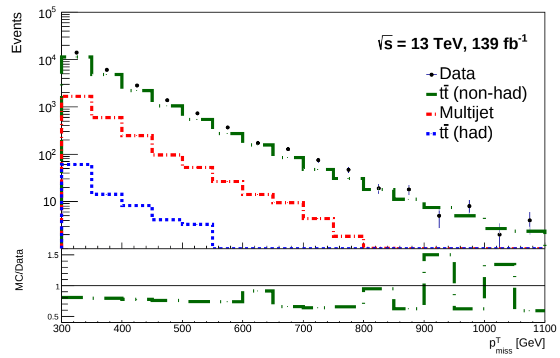

Figure 4.7: Histogram showing the magnitude of the missing transverse momentum in each event for the various SM processes at reco level. The $t\bar{t}$ (non-had) in green has the leading contribution, followed by multijet in red and $t\bar{t}$ (had) in blue. The ratio between leading SM contribution and data is shown in the ratio plot.

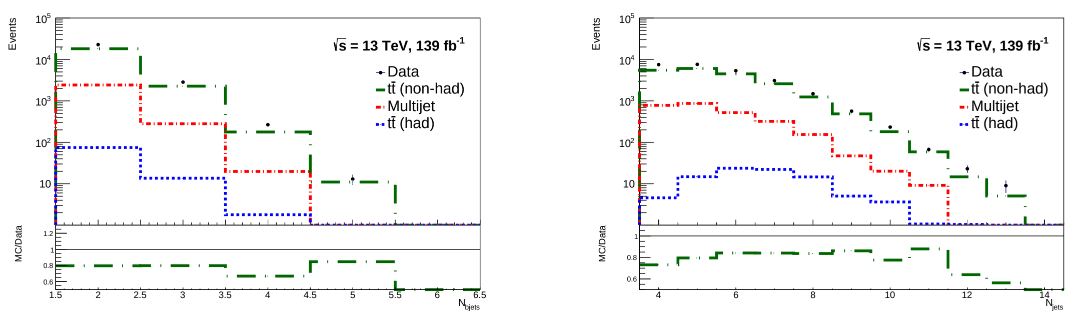

Figure 4.8: Histogram showing the number of b-tagged jets in each event for the various SM processes at reco level on the right-hand side and the number of jets in each event on the left-hand side. The $t\bar{t}$ (non-had) in green has the leading contribution followed by multijet in red and $t\bar{t}$ (had) in blue. The ratio between leading SM contribution and data is shown in the ratio plot in both distributions.

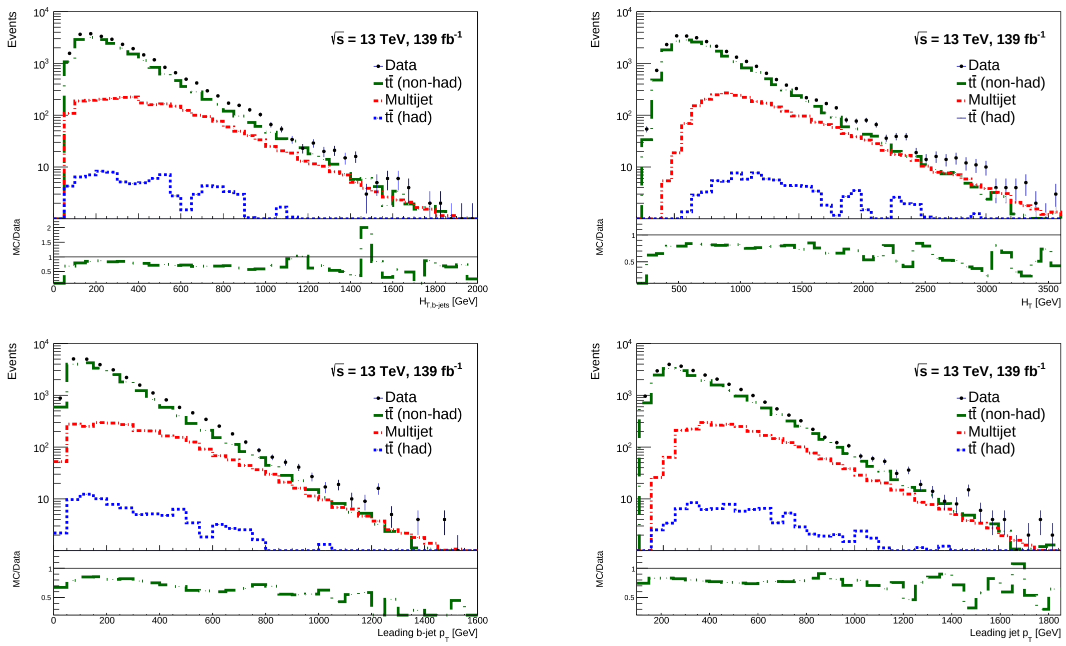

Figure 4.9: Histograms on the top panel showing the H_T of all jets and $H_{T,bjets}$ of b-jets in each event for the various SM processes at reco level. Histograms on the bottom panel showing the leading jet p_{T} and leading b-jet p_{T} in each event for the various SM processes at detector-level. The $t\bar{t}$ (non-had) in green has the leading contribution, followed by multijet in red and $t\bar{t}$ (had) in blue. The ratio between leading SM contribution and data is shown in the ratio plot of each histogram shown

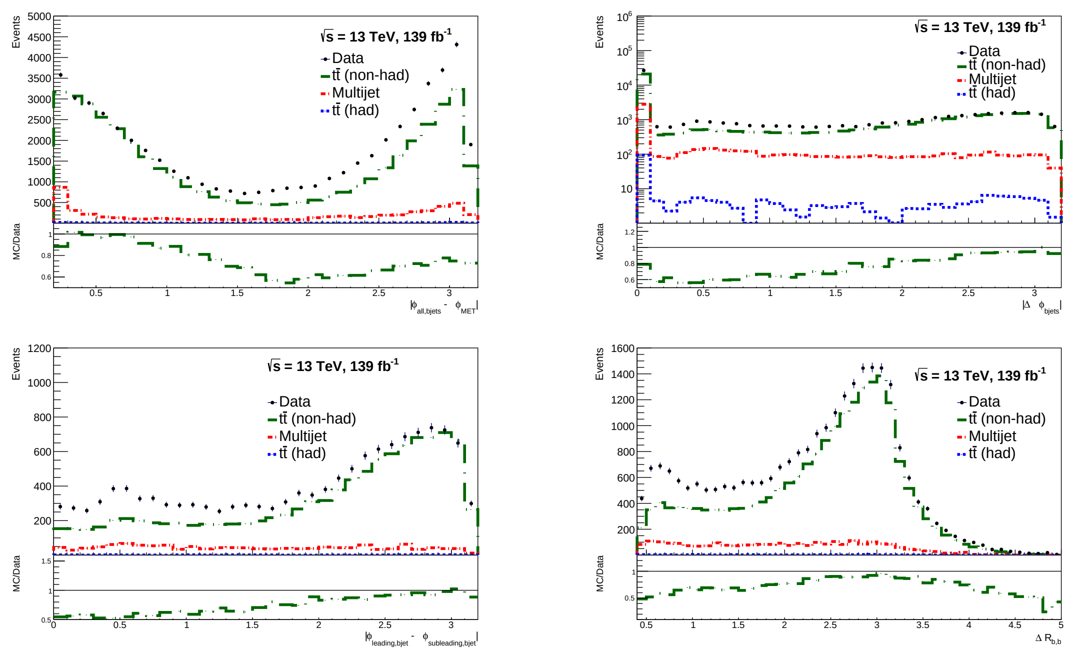

Figure 4.10: Histograms on the top panel showing the $|\phi_{all,bjets} - \phi_{MET}|$ and $\Delta\phi_{bjets}$ in each event for the various SM processes at reco level. Histograms on the bottom panel showing $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$ and $\Delta R_{b,b}$ in each event for the various SM processes at detector-level. The $t\bar{t}$ (non-had) in green has the leading contribution, followed by multijet in red and $t\bar{t}$ (had) in blue. The ratio between leading SM contribution and data is shown in the ratio plot of each histogram shown

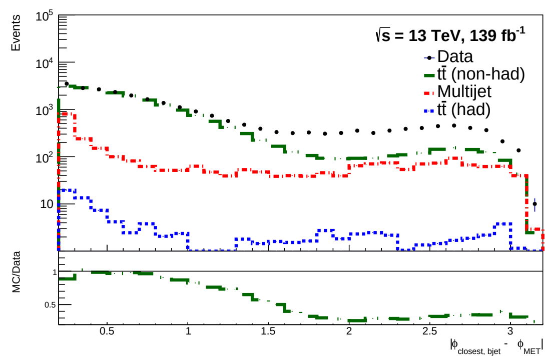

Figure 4.11: Histogram showing the $|\phi_{closest,bjet} - \phi_{MET}|$ in each event for the various SM processes at reco level. The $t\bar{t}$ (non-had) in green has the leading contribution, followed by multijet in red and $t\bar{t}$ (had) in blue. The ratio between leading SM contribution and data is shown in the ratio plot of each histogram shown

Since there was a distinct leading SM process in Figure 4.7 to Figure 4.11, the unfolding was performed using $t\bar{t}$ (non had) PYTHIA and SHERPA MC samples. It is worth noting that other leading backgrounds, such as multijet and $t\bar{t}$ (all-had), were also plotted but were orders of magnitude smaller compared to $t\bar{t}$ (non-had). In Figures 4.12 - 4.23, the distributions on the left-hand side are unfolded using PYTHIA MC samples and on the right-hand side using SHERPA. The distributions which depict the unfolded data all confirm and agree with trends discussed for the particle-level and uncorrected data distributions shown above.

Looking at the data closely, the unfolded data is in agreement with the theoretical predictions for both PYTHIA and SHERPA, with the unfolded data yielding more events as predicted by the SM. In Figure 4.23, there seems to be an interesting deviation that happens at $|\phi_{closest,bjet} - \phi_{MET}| > 1.6$; the particle-level events exceed the detector-level and unfolded data events. This is interesting because this is not expected, and the reason behind this deviation will have to be investigated. It is worth keeping in mind that the IB technique contributes to some uncertainties in the unfolding. As mentioned, it has an incomplete regularisation, meaning the unfolded data will have some statistical fluctuations. Any form of uncertainty in the data will be further propagated in this technique.

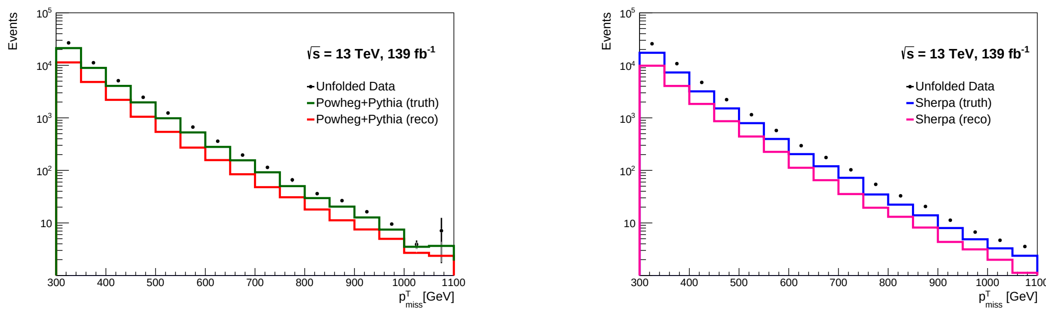

Figure 4.12: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the magnitude of the missing transverse momentum in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

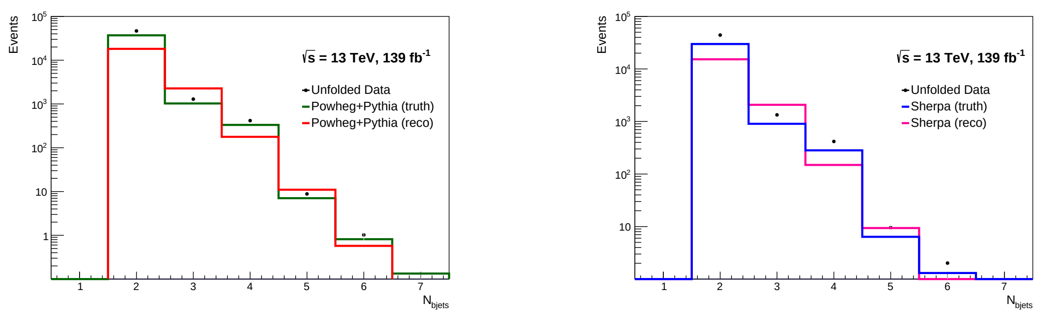

Figure 4.13: Histograms showing the unfolded data alongsidePYTHIA and SHERPA MC samples at particle- and detector-level measuring the number of b-tagged jets in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

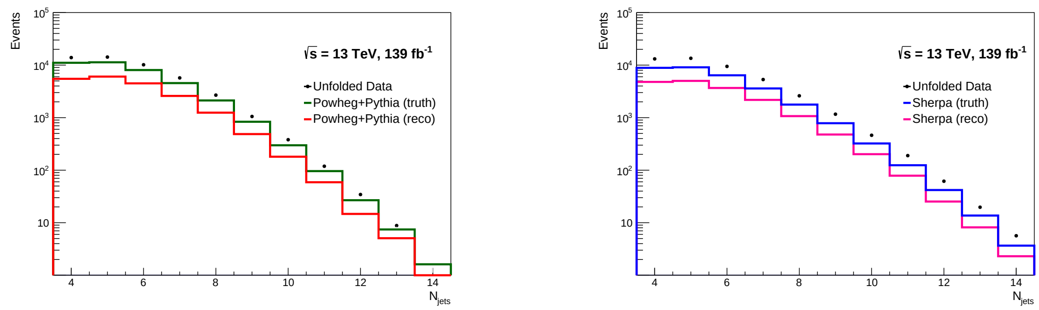

Figure 4.14: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the number of jets in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

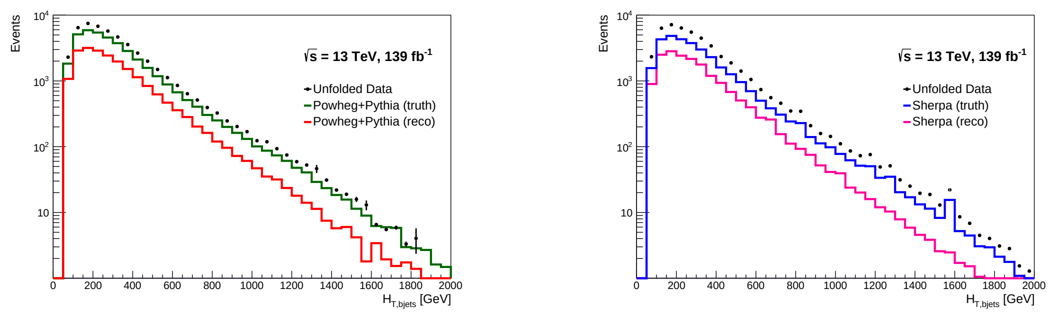

Figure 4.15: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the scalar sum p_{T} known as $H_{T,bjets}$ of b-tagged jets in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

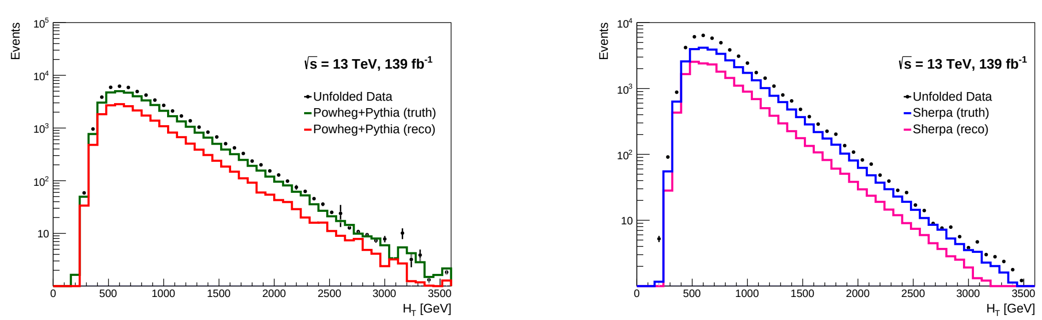

Figure 4.16: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the scalar sum p_{T}, known as H_T of all jets in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

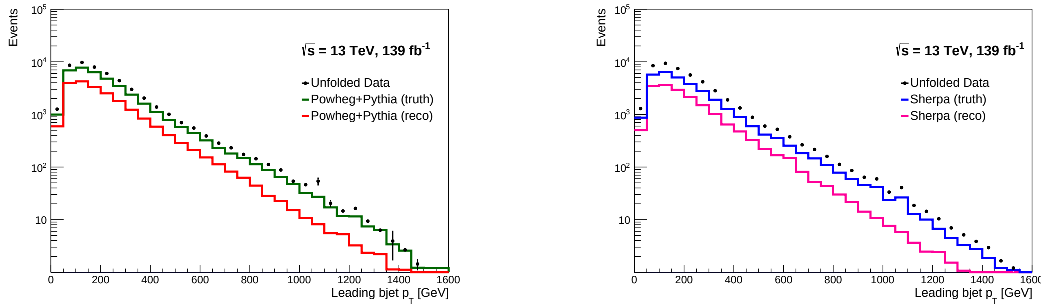

Figure 4.17: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the leading b-jet p_{T} in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

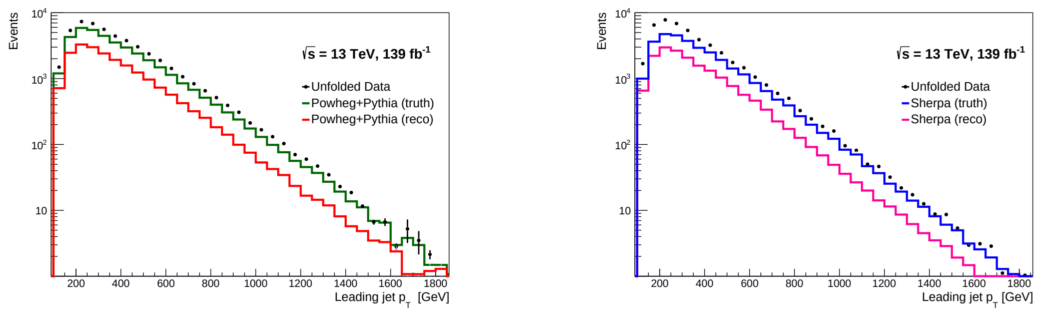

Figure 4.18: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the leading jet p_{T} in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

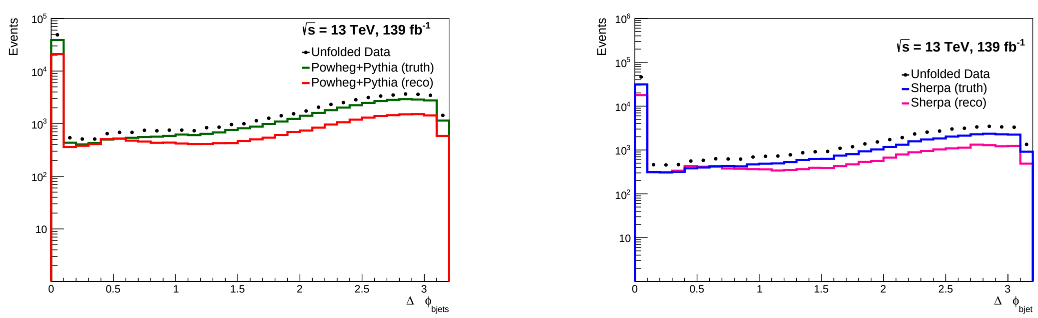

Figure 4.19: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the $\Delta\phi_{bjets}$ in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

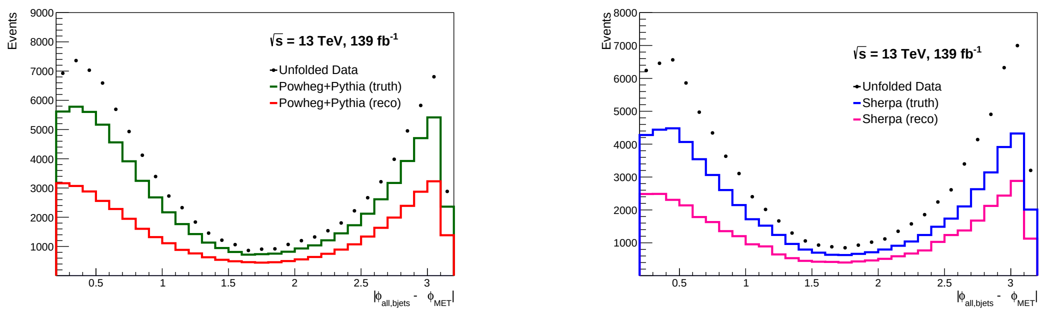

Figure 4.20: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the $|\phi_{all,bjets} - \phi_{MET}|$ in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

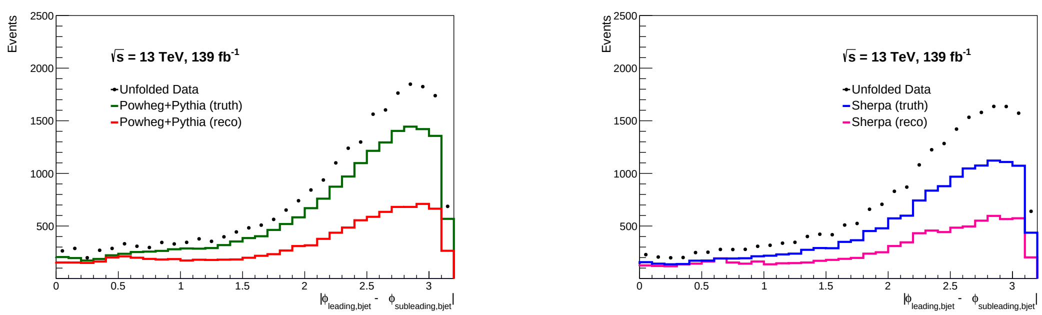

Figure 4.21: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$ in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

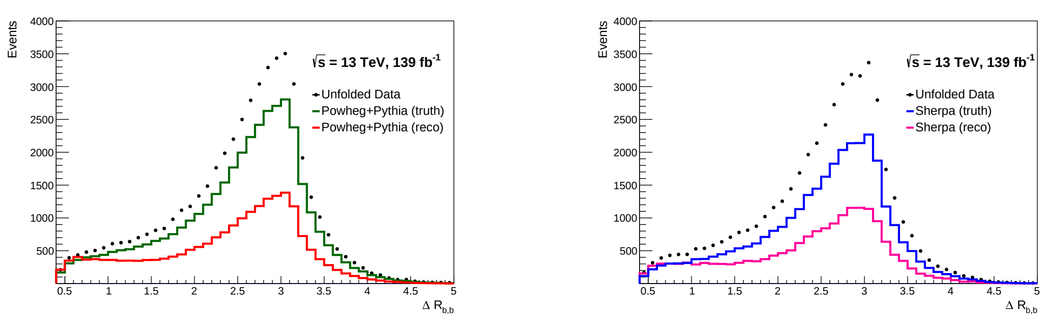

Figure 4.22: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring the $\Delta R_{b,b}$ in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

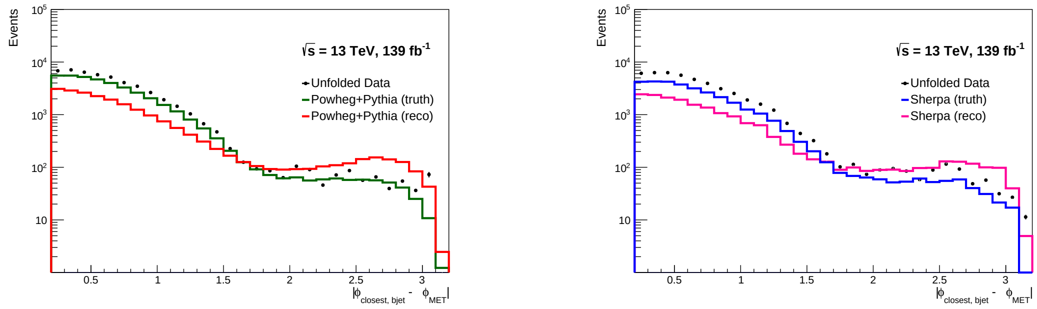

Figure 4.23: Histograms showing the unfolded data alongside PYTHIA and SHERPA MC samples at particle- and detector-level measuring $|\phi_{closest,bjet} - \phi_{MET}|$ in each event. The distribution on the left-hand side is unfolding with PYTHIA whilst the distribution on the right-hand side is unfolding SHERPA.

In Figure 4.24 to Figure 4.28, the unfolding comparison between the two MC samples is presented. The distribution all exhibit a similar trend; the unfolded PYTHIA and SHERPA data are equal, or the PYTHIA unfolded data is slightly higher; this is observed in the ratio plot in the distributions. As mentioned, the differences between the unfolded data are due to systematic uncertainties and differences in the modelling of the MC event generators used. In theory, the two unfolded samples are meant to have the same results.

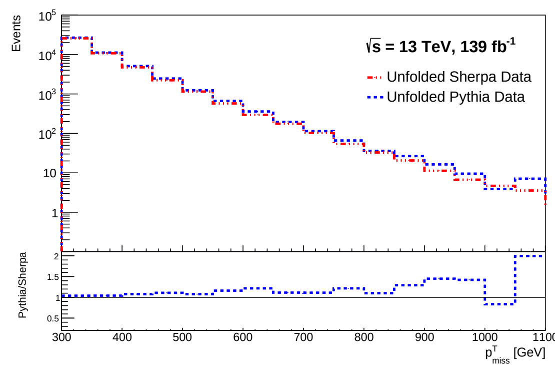

Figure 4.24: Histogram showing the unfolding results for the magnitude of the missing transverse momentum in each event using PYTHIA and SHERPA MC samples. The red is the unfolding result for using SHERPA and the blue is the unfolding result using PYTHIA. The ratio between the two unfolded data is shown in the ratio plot.

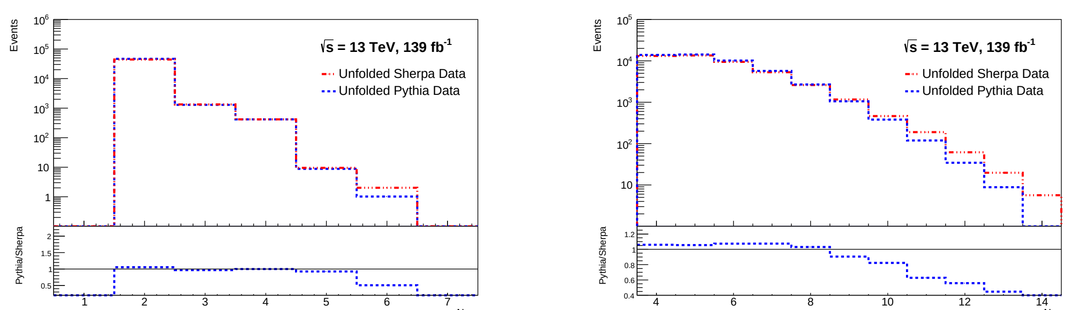

Figure 4.25: Histogram showing the unfolding results for the number of b-tagged jets (left) and the number of jets (right) in each event using PYTHIA and SHERPA MC samples. The red is the unfolding result for using SHERPA and the blue is the unfolding result using PYTHIA. The ratio between the two unfolded data is shown in the ratio plot.

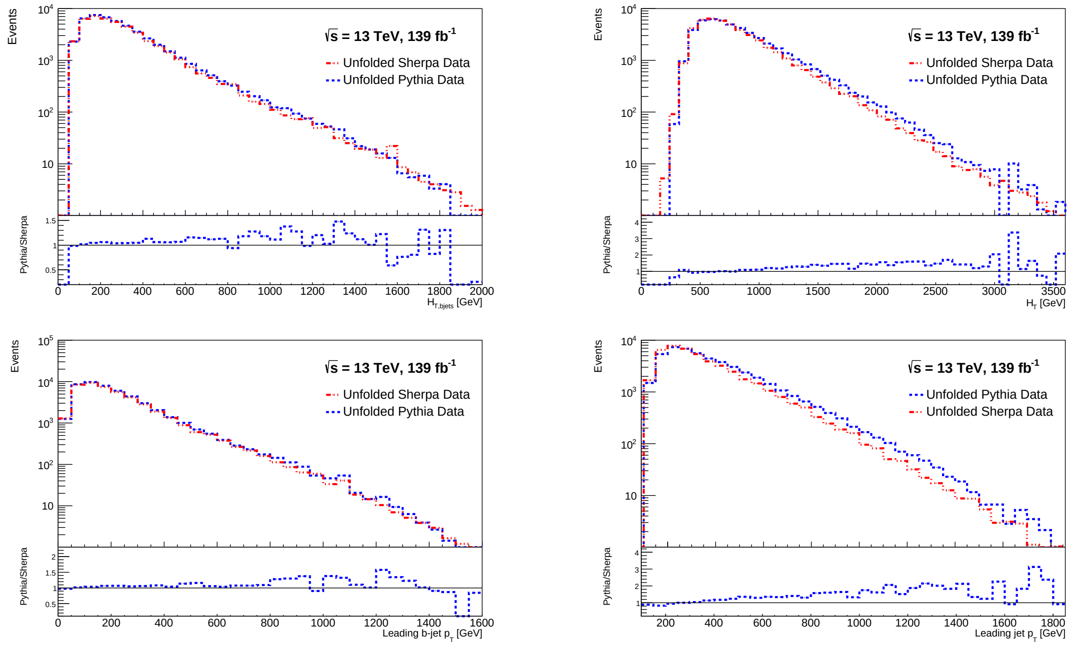

Figure 4.26: Histogram showing the unfolding results for leading b-jet p_{T}, leading jet p_{T}, H_T and $H_{T,bjets}$ in each event using PYTHIA and SHERPA MC samples. The red is the unfolding result for using SHERPA and the blue is the unfolding result using PYTHIA. The ratio between the two unfolded data is shown in the ratio plot.

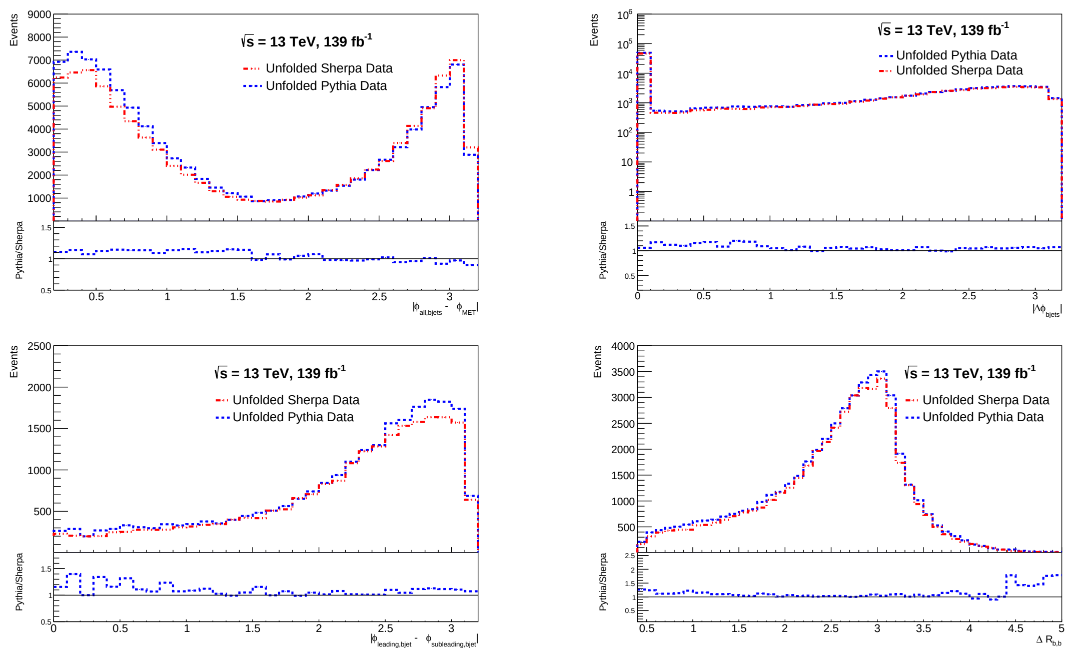

Figure 4.27: Histogram showing the unfolding results for $|\phi_{all,bjets} - \phi MET|$, $|\Delta\phi_{bjets}|$, $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$ and $\Delta R_{b,b}$ in each event using PYTHIA and SHERPA MC samples. The red is the unfolding result for using SHERPA and the blue is the unfolding result using PYTHIA. The ratio between the two unfolded data is shown in the ratio plot.

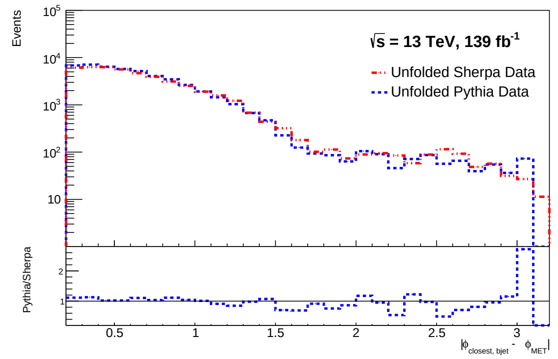

Figure 4.28: Histogram showing the unfolding results for $|\phi_{closest,bjet} - \phi_{MET}|$ in each event using `PYTHIA` and `SHERPA` MC samples. The red is the unfolding result for using `SHERPA` and the blue is the unfolding result using `PYTHIA`. The ratio between the two unfolded data is shown in the ratio plot.

In the distributions shown thus far, the data unfolded using `PYTHIA` MC samples is in better agreement with theoretical predictions. In Figure 4.29 and Figure 4.30, the unfolded data using `PYTHIA` is presented and plotted against detector-level events for E_T^{miss} and $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$. In both distributions the `PYTHIA` event samples are in better agreement with the unfolded data.

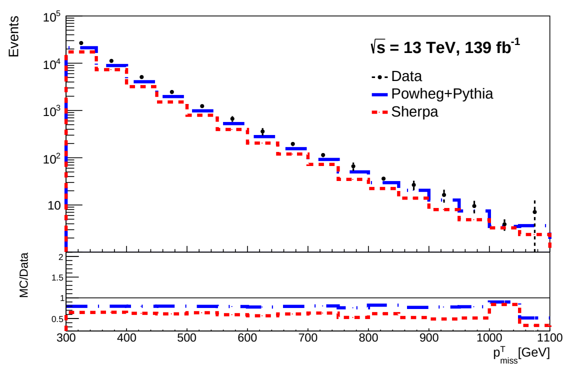

Figure 4.29: Distribution showing the `PYTHIA` unfolded data alongside `SHERPA` MC events at detector-level and `PYTHIA` events at detector-level measuring E_T^{miss} in each event. The ratio between the unfolded data and the different MC event generators is also depicted.

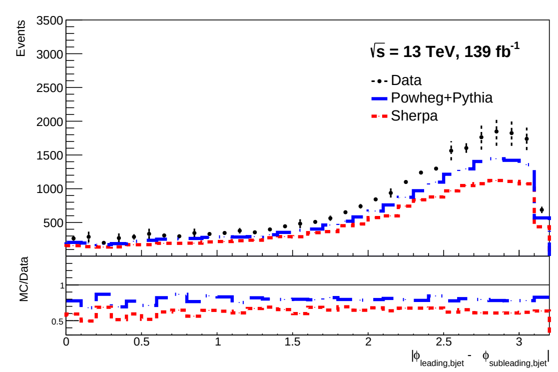

Figure 4.30: Distribution showing the `PYTHIA` unfolded data alongside `SHERPA` MC events at detector-level and `PYTHIA` events at detector-level measuring $|\phi_{leading,bjet} - \phi_{subleading,bjet}|$ in each event. The ratio between the unfolded data and the different MC event generators is also depicted.

4.8 Conclusion

The results presented are for the measurement of the b-jet cross-section using Run 2 ATLAS data collected at a centre-of-mass $\sqrt{s} = 13$ TeV at an integrated luminosity of 139 fb^{-1}. The measured observables were corrected for detector effects and limited resolution using the RooUnfold framework so they could be directly compared with theoretical distributions. Initially, the uncorrected data and particle-level events were plotted to ensure that an SM process was not excluded or unaccounted for in the measurement. The results confirmed that $t\bar{t}$ with a non-all hadronic final state has the leading SM contribution, so the measurement was done using $t\bar{t}$ samples only. The unfolding was performed using two MC samples, namely, `SHERPA` and `PYTHIA`; a comparison was made between MC event generators. The difference in the unfolding of the two MC event generators is due to systematic uncertainties. It was observed that the `PYTHIA` event generator was in better agreement with the theoretical prediction.

Chapter 5

Performance Studies

This Chapter explores performance studies with ATLAS and ALICE detectors at the LHC in CERN. Performance studies are also an important aspect in experimental particle physics, as they are meant to test and optimise object reconstruction, such as tracking and jet reconstruction algorithms, ensuring the precision and reliability of the experimental set-up. The first performance study presented looks at b-tagging in ATLAS. There has been noticeable growth in the use of machine learning in heavy-flavour tagging. From the use of boosted decision trees to all-in-one transformers, which will be discussed. The second performance study in this Chapter deals with b-tagging in ALICE. The ALICE experiment is specifically designed to study a state of matter known as the quark-gluon plasma, which is believed to have existed in the early universe [124]. The development of a b-tagging algorithm in ALICE will be discussed. ALICE has not previously focused on b-jet identification, making heavy-flavour tagging a promising new area of exploration. The development presents significant opportunities to use heavy-flavour jets to deepen the understanding of jet physics within the experiment.

5.1 DESY Summer Programme: Flavour Tagging Performance Studies with ATLAS at the LHC

5.1.1 Heavy-flavour Tagging Machine Learning Algorithms

As mentioned, HF tagging exploits the decay properties of b- and c-hadrons. The search for HF jets involves looking for displaced vertices, among other properties. The other properties are a large track impact parameter d_0, missing hits on inner detector layers, semi-leptonic decays, and hard fragmentation. The impact parameter can be seen in Figure 1.2 as d_0, which is as a result of the secondary vertex. Missing hits on the detector and semi-leptonic decay help to easily identify interesting events whilst suppressing unwanted backgrounds. One of the decay channels for $t\bar{t}$ is semi-leptonic decay; from the name, it decays into b-quarks and leptons.

Each step in the growth of ML algorithms in ATLAS, from the first versions developed during Run 1, which was centred around likelihood-based discriminant and boosted decision trees, to recurrent and deep neural networks, is marked by significant improvements over the previous years. This has led to the development of GN2, which is based on GN1.

The current approach to flavour-tagging in ATLAS is a two-level tagger [125]. It uses hadronic jets and charged particle jet tracks as inputs to execute and achieve jet classification [125]. The low-level taggers are known as Recurrent Neural Network (RNNIP), Deep-Sets (DIPS), Secondary Vertex (SV1), JetFitter, and IPxD [52]. These can further be separated into manually optimised and track-based taggers. The track-based taggers, RNNIP and DIPS, use the track input information and train on tracks a neural network. The manually optimised taggers (IPxD, SV1 and JetFitter) exploit the decay properties of b- and c-hadrons. IPxD is also a track-based tagger that reconstructs the impact parameter [125]; SV1 is a vertex tagger and reconstructs the secondary vertex, and JetFitter [125] attempts to form the decay chain of the vertices using final state products. The outputs of the low-level taggers are fed into a deep-learning classifier (DL1), which is the high-level tagger as inputs. This is illustrated in Figure 5.1. The DL1 series, as mentioned, is a multivariate algorithm that takes the outputs of the low-level tagger as inputs and attempts to achieve maximum classification performance. DL1d, which is the recommended tagger for Run 3, is similar to DL1r; the difference is the upgraded RNNIP using Deep-Sets [125]. The output from the DL1 series is jet flavour predictions: p_c, p_b, and p_u. There have been some challenges with the DL1 series, such as difficulty and complexity when working with reconstructed jets. The tuning for different use cases requires lots of single steps before the final tagger, and it creates a dependence on the low-level taggers. If one of the low-level taggers does not perform well, it ultimately affects the results obtained at the end. These challenges meant that there was room for improvement for the DL1 series.

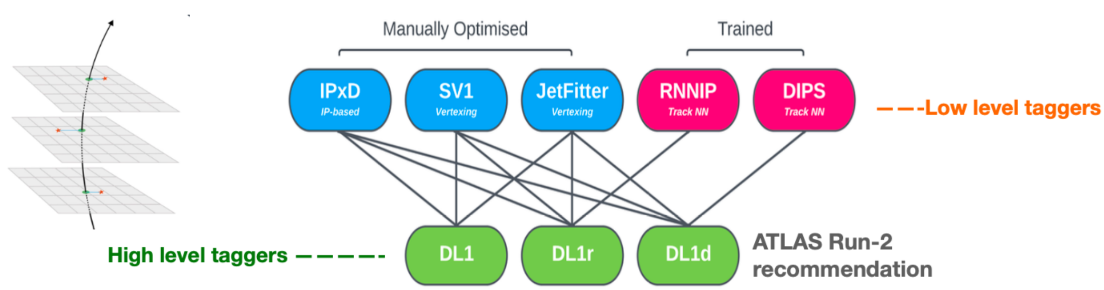

Figure 5.1: Diagram illustrating the two-tier tagging approach in ATLAS collaboration. Low-level consisting of trained and manually optimised tagger and the DL1 series of high-level taggers taken from [126].

GN2, as briefly mentioned, provides an innovative approach to the recognised two-level tagger approach in ATLAS. It works similarly to the manually optimised taggers in low-level but with a unified architecture. It aims at fulfilling two goals: (1) grouping tracks that originate from the same vertex and (2) using the information from the grouped tracks to predict the underlying physics [125]. GN2 is an upgraded version of GN1 [125], which is one of the first improvements from DL1d. The Graph

Attention Network in GN1 is replaced with a transformer in GN2. The track information, such as track parameters, uncertainties, impact parameters, and detailed hit information, alongside jet kinematics such as p_{T} and η, are directly fed into the transformer with multi-head attention layers. There are three ML tasks that are trained simultaneously: (1) jet flavour predictions, (2) vertexing, and (3) track origin predictions. The set-up for GN2 can be easily optimised for various cases and track and/or jet improvements. The more refined details about layers in the transformer encoder, the number of attention heads and dimensions can be found in 5.1.2. With the new setup of GN2, it can also predict underlying physics for tau-jets. The unified architecture is illustrated in Figure 5.2.

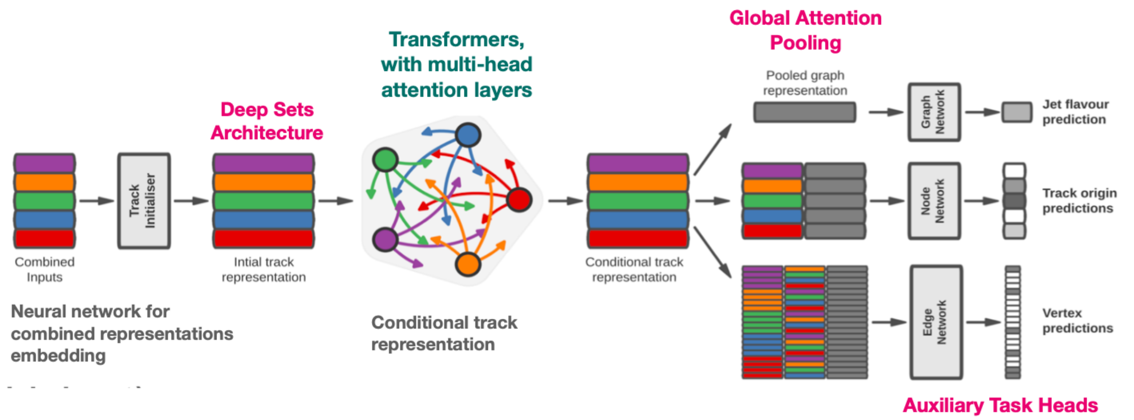

Figure 5.2: A illustration of the GN2 architecture taken from [127]

Before the inputs are fed into the transformer, they need to be preprocessed. This involves resampling the jet p_{T} and η for each flavour. When plotting the jet p_{T} and η for b-, c-, tau-, and light-jets, the number of jets for each p_{T} value would vary as the underlying physics for each jet is different, as seen in Figure 5.3. The concept of resampling involves applying weights to the events so that the p_{T} and η distributions for different jet flavours appear similar. This is implemented to avoid any biases, where one flavour could be trained better because it peaks first in the distribution. Once the jet kinematics are resampled, the distributions would appear similar to Figure 5.4. Rather than providing a form of discrimination for the training, the jet kinematics and track information act as parameters. In Table 5.1 is a list of all other inputs that can be used in the GN2 transformer.

!

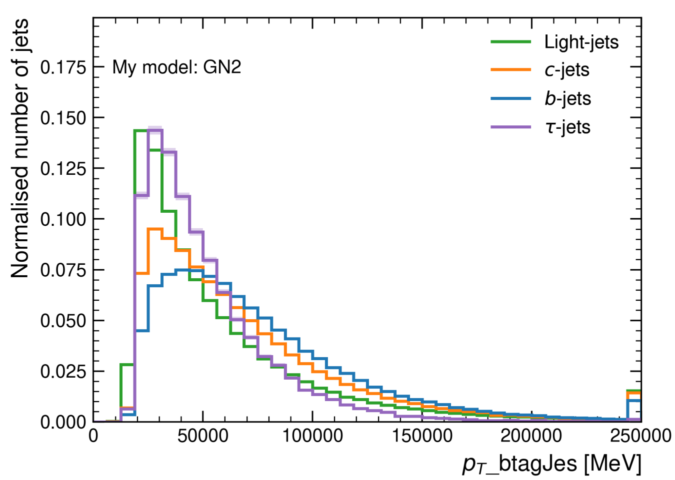

Figure 5.3: Distribution of normalised number of jets as a function jet p_{T}. This is the distribution of the jet kinematics before resampling.

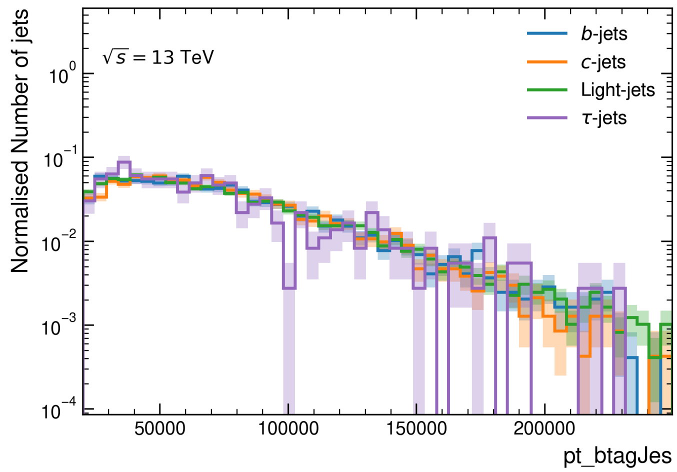

Figure 5.4: Distribution of normalised number of jets as a function jet p_{T}. This is the distribution of the jet kinematics after resampling.

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)
dη	Pseudorapidity of the track, relative to the jet η
dϕ	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the longitudinal plane
$z_0\sin\theta$	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
$s(z_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used in te reconstruction of an electron or muon (only for GN1 Lep)

Table 5.1: List of various inputs used in the GN2 transformer taken from [127]

5.1.2 Optimisation Strategy

The main idea behind the performance study was to investigate if the GN2 training could be optimised by (1) using different resampling compositions and statistics, (2) using inputs that have split c-jet to exploit decay properties of C-hadrons and (3) using an expanded GN2 algorithm network. The ROC curve and b-jet efficiency vs the background rejection were then plotted to gauge the performance of the GN2 network. The details of the jet flavour composition used as an input in each optimisation can be found in Chapter 5.1.3.

Initially, before the training was done, the different jet kinematics (jet p_T and η) were plotted before and after resampling for different jet flavours, i.e., b-jet, c-jet, light-jet, and tau-jets, to obtain a visual representation of what resampling is. Figure 5.3 shows jet p_T before resampling, and Figure 5.4 shows the same jet distribution but after resampling.

The optimisation of the training using different resampling and statistics entailed increasing the number of input jets in the training, and for the GN2 default with 100 million jets, the new sampling strategy involves doubling the number of light jets rather than maintaining the same number as before. As a result, during the GN2 training, the weight applied to light jets would be twice that of b- and c-jets. The second training attempted to optimise the results by using inputs that exploit decay properties of c-hadrons. Introducing the c-split means instead of having p_c, there would be p_{D0} (D^0 meson) and p_{D0} (non-D^0 meson). The motivation behind the c-split is that c-jets are tagged by looking for D^0 and D^+ mesons, as shown in Table 5.2, and there's a significant difference in the lifetimes of the charged and neutral D mesons; the lifetime of a charged D meson is six times longer than the lifetime of a neutral D meson, which could impact the GN2 network. The last training was to see if expanding the GN2 algorithm network would improve the performance..

Charm Particle	Quark content
D^+	$c\bar{d}$
D^0	$c\bar{u}$
F^+	$c\bar{s}$
Λ_c^+	cud

Table 5.2: List of charm particles used for c-tagging and the quark content in each of the charm particles taken from [128]

5.1.3 Datasets

The MC event samples used in the flavour tagging performance study used both Run 2 and Run 3 $t\bar{t}$ MC samples. The training was executed using the `SALT` framework [129] which is part of flavour-tagging tools in ATLAS and the plotting of the distribution was done using `PUMA-hep` [130]. The configuration file used for the machine-learning GN2 algorithm was provided in the `SALT` framework [129]. The MC event samples used in the training were generated using `UMAMI` [131] with increasing statistics to to see the effect on the performance. The composition of the samples I used to investigate the dependence on statistics is as follows:

- GN2 default with 30 million jets: 9 million b-jets, 9 million c-jets, 9 million light-jets, 3 million τ-jets.
- GN2 default with 100 million jets and new sampling: 25 million b-jets, 25 million c-jets, 50 million light-jets, 4 million τ-jets.

To investigate what effect introducing a new discriminatory factor into the training will have on the performance, a new set of samples were generated using `UMAMI`. New samples with c-split were used in the training, the composition of the samples used was as follows:

- GN2 with c-jet split and 100 million jets: 25 million b-jets, 12.5 million D0 meson c-jets, 12.5 million non-D0 mesons c-jets, 50 million light-jets, 4 million τ-jets.
- GN2 with c-jet split and 30 million jets: 7.2 million b-jets, 3.6 million D0 meson c-jets, 3.6 million non-D0 meson c-jets, 14.4 million light-jets, 1.2 million τ-jets

To investigate if expanding the GN2 machine learning algorithm network would improve the performance, the layers in the GN2 configuration file provided by `SALT` was increased from 6 to 8 and an additional hidden layer was added to the network to expand it. The composition of the new samples used in the training was a follows:

- GN2 with c-jet split and 100 million jets: 25 million b-jets, 12.5 million D0 meson c-jets, 12.5 million non-D0 meson c-jets, 50 million light-jets, 4 million τ-jets, 8 layers and hidden layer of [128, 64, 32, 16]

5.1.4 Constructing The Discriminant

As previously mentioned and illustrated in Figure 5.2, the GN2 transformer has three auxiliary tasks; the optimisation focuses on jet flavour prediction: p_c, p_b, and p_u. Using the jet flavour probabilities, the b-jet discriminant was computed. The b-jet discriminant is one of the variables that can be calculated and plotted to evaluate the performance of the training. It is calculated using the following equation (5.1):

$$D_b = log\left[\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_u}\right] \tag{5.1}$$

The values for f_c and 1-f_c are free parameters which trade off between background rejections such that the larger f_c equates to more c-jet rejection. Similarly to the value of the probabilities, where the sum of the jet flavour predictions is 1, the sum of f_c and 1 f_c also has to be 1. The values for these free parameters are calculated by optimising the c-jet and light-jet rejection. In equation (5.1), f_c and f_u act as weights being applied to the jet flavour probabilities. With the recent GN2 setup, you can also calculate the b-tag discriminant using tau class flavour prediction, and the equation would look like equation (5.2):

$$D_b = log\left[\frac{p_b}{f_c \cdot p_c + f_\tau \cdot p_\tau + (1 - f_c - f_\tau) \cdot p_u}\right] \tag{5.2}$$

For GN2, it is set to $f_c = 0.2$ and $f_\tau = 0.02$, which has already been used for GN2 default in other studies. In Figure 5.5, as shown, the b-jet discriminant is plotted against the normalised number of jets; ideally, there has to be a clear separation between b-jets and non-b-jets. Figure 5.5 is the discriminants plotted for one of the optimisations. It demonstrates that the b-jets have been identified to be different from the other jets, which is the main objective. The five %'s shown on the distribution are called working points (WPs), which are critical for calibrating the performance of the tagging algorithm. Each WP represents a trade-off between the efficiency of tagging b-jets and the background rejection. In other words, they are calibration points that help in adjusting the balance between correctly identifying b-jets and minimising false positives (incorrectly tagging non-b-jets as b-jets). These WPs correspond to different b-jet efficiencies, which will be further analysed in the ROC curve distribution.

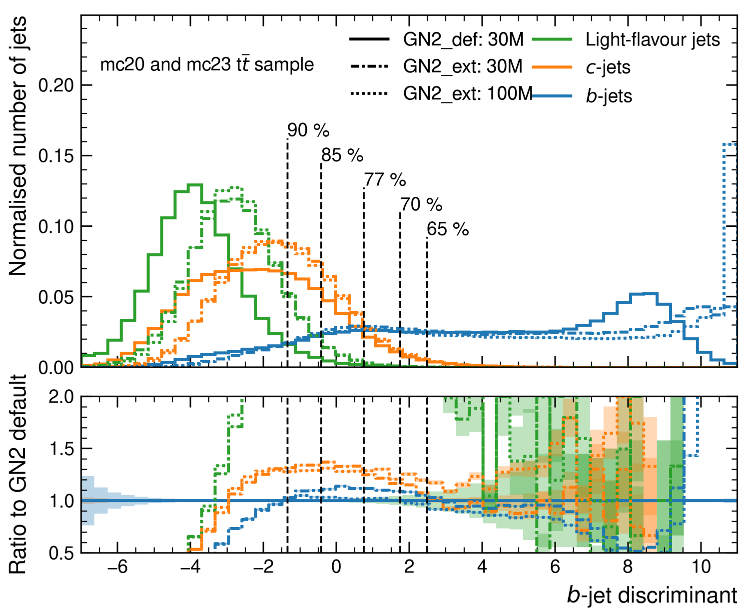

Figure 5.5: The %'s are called working points and are used for calibration. They relate to the b-jet efficiencies seen in ROC curves. In the distribution, ideally, there should be a clear separation between b-jets and non b-jets.

5.1.5 Results

In addition to constructing the discriminant, the optimisation can also be measured by plotting b-jet efficiency and comparing the training in a ROC curve. An ROC curve displays the relationship between background rejection and b-jet efficiency. The performance of a flavour-tagging algorithm is characterised by the probability of correctly tagging a signal jet, known as efficiency, ε, and the probability of mistakenly identifying a background jet; this is known as the mis-tag rate. When the identification of b-jets is required, the tagging efficiency of a given requirement on the b-tagging discriminant is denoted by ε_b, and the charm-jet and light-flavour jet rejection factors are $1/\varepsilon_c$ and $1/\varepsilon_{light}$.

In Figure 5.6, a comparison between GN2 default with 30 million jets and GN2 default with 100 million jets and new sampling is presented. The comparison is meant to explore if the GN2 algorithm is statistic dependent. The b-jet efficiency is plotted from a value of 0.5, which is 50% accuracy in identifying b-jets. Lower values are not plotted, as it is not useful to investigate accuracies below 50%. The values for f_c and f_τ are as shown on Table 5.3; these values are taken from previous studies [125]. In Figure 5.6, at lower efficiencies, there is a higher background rejection, and what this means is at 50% b-tagging accuracy, more c- and light-jets are rejected. On the other hand, at 100% b-tagging accuracy, fewer c- and light-jets are rejected. The working point 70% is commonly used because it maintains a good balance between a high background rejection and the b-jet efficiency value, which is relatively close to 100%.

In the top panel of Figure 5.6, as mentioned, at a lower efficiency more c- and light-jets are rejected, and as the b-jet efficiency increases, fewer c- and light-jets are rejected. In the ratio panels, the performance is measured with respect to the GN2 default with 30 million jets. In the light-jet ratio panel of Figure 5.6, the GN2 with 100 million jets and new sampling at 70% efficiency performs approximately 1.15 times better at identifying and rejecting light-jets compared to the GN2 default with 30 million jets. In the c-jet ratio panel, the GN2 with 100 million jets and new sampling in Figure 5.6 at 70% efficiency performed close to 1.2 times better at identifying c-jets and rejecting them.

Table 5.3: The f_c and f_τ values in calculating the discriminant and b-jet efficiency taken from [125]

	GN2
f_c	0.2
f_τ	0.02

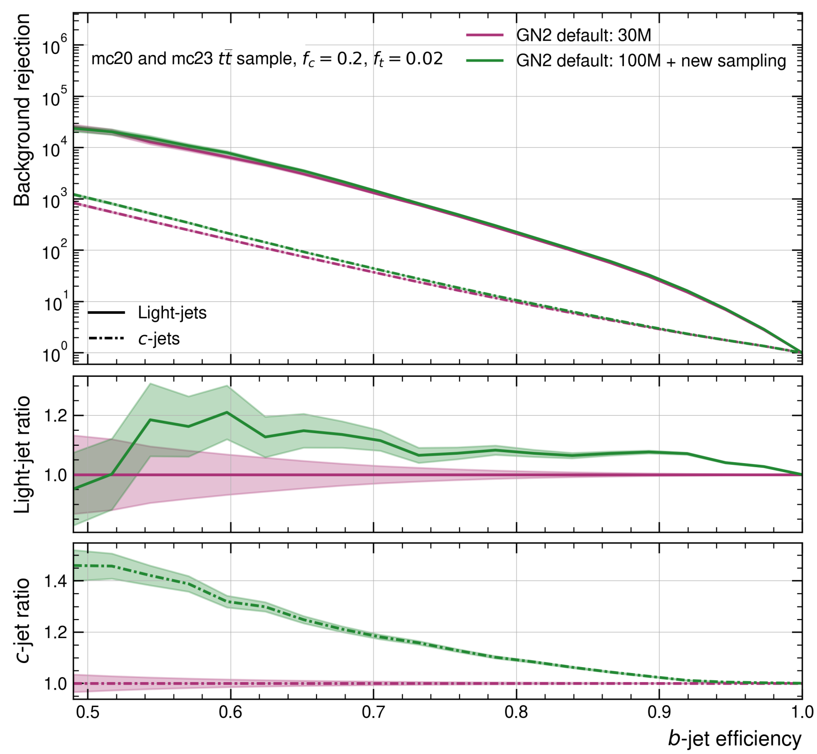

Figure 5.6: The c-jet and light-flavour jet s as a function of the b-jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_{\mathrm{T}} < 250$ GeV. The ratio with respect to the performance of the GN2 default with 30 million jets is shown in the bottom panels. The error bands are denoted by the shaded regions.

In Figure 5.7, the comparison is meant to explore if introducing a c-split between the c-hadrons would result in better performance of GN2 set-up. The motivation behind this separation has been previously explained.

In Figure 5.7, a comparison between GN2 default with 100 million jets with new sampling against GN2 with 30 million jets and c-split alongside the GN2 set-up with 100 million and c-split is shown. As mentioned and observed in Figure 5.6, at a lower efficiency more c- and light-jets are rejected, and as the b-jet efficiency increases, fewer c- and light-jets are rejected for all three GN2 set-ups. In the ratio panels, the performance is measured with respect to the GN2 default with 100 million jets with new sampling. In the light-jet ratio panel, the GN2 with 100 million jets and c-split at 70% efficiency performs close to 0.3 times less at identifying light-jets and rejecting than GN2 with 30 million and c-split, which performs 0.45 times less. In the c-jet ratio panel, the GN2 with 100 million jets and c-split at 70% efficiency performed close to 1.08 times better at identifying c-jets and rejecting them, and for GN2 with 30 million jets and c-split, it performed close to 0.04 times less than the default. From observation, there is an implicit trade-off between light- and c-jets; this could originally be the light-jets being mis-tagged. This informs us that there is room for improvement when it comes to the GN2 network. The GN2 with 100 million jets and c-split performing better than the GN2 with 30 million jets and c-split confirms that performance is indeed dependent on statistics. The more jets in training, the better performing the GN2 set-up.

In Figure 5.8, the comparison is meant to explore if expanding the network would result in better performance of the GN2 setup. The details on how the network was expanded can be found in Chapter 5.1.2. The motivation behind this separation has been previously explained. When comparing GN2 with 100 million jets and new sampling with GN2 100 million jets and c-split expanded with 8 layers and an additional hidden layer. In Figure 5.7, it shows that introducing the c-split decreased the performance for light-jets, so by comparing the set-ups with the same number of jets, a direct comparison can be made to understand if expanding the network could improve the identification and rejection of light-jets. As mentioned and observed in figure 5.6, at a lower efficiency more c- and

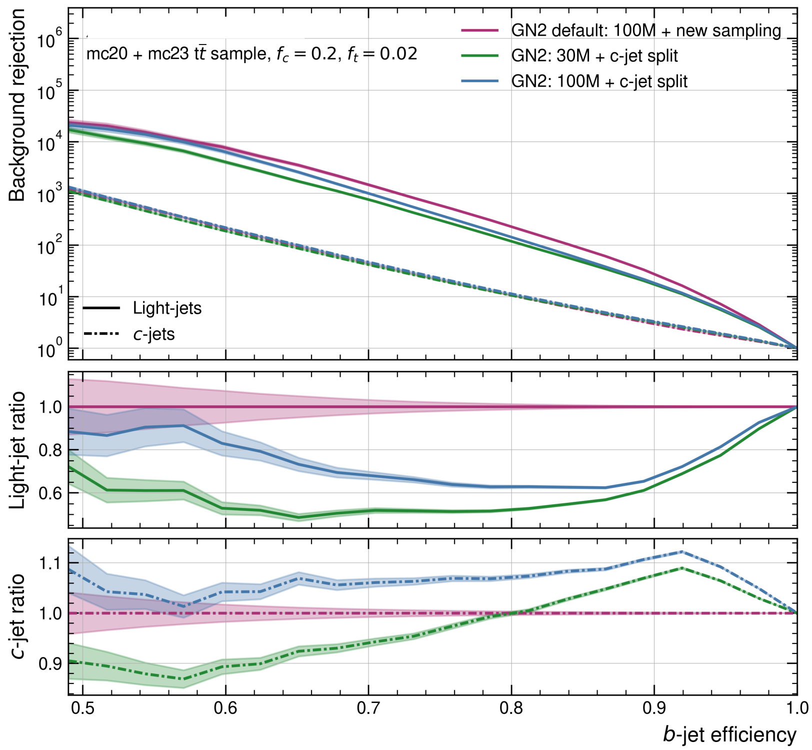

Figure 5.7: The c-jet, light-flavour jet, and b-jet rejections as a function of the b-jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_{\mathrm{T}} < 250$ GeV comparing three GN2 set-ups. With a different number of jets and introducing a c-split. The error bands are denoted by the shaded regions.

light-jets are rejected, and as the b-jet efficiency increases, fewer c- and light-jets are rejected for both GN2 setups.

In the ratio panels, the performance is measured with respect to the GN2 with 100 million jets. In the light-jet ratio panel, the GN2 with 100 million jets and GN2 with 100 million, c-split and expanded network at 70% efficiency performed relatively the same at identifying light-jets and rejecting them. When comparing this result with Figure 5.7, it is observed that expanding the network does improve the performance with respect to light jets. In the c-jet ratio panel, the GN2 with 100 million with c-split and expanded network at 70% efficiency performed close to 0.2 times less at identifying c-jets and rejecting them and for GN2 with 100 million jets, which is the default. When comparing this result with Figure 5.7, expanding the network does not improve the performance with respect to c-jets. From observation, there is another implicit unexpected trade-off between light- and c-jets. Introducing c-split improves the performance for c-jet rejection but decreases the performance for light-jet rejection. On the other hand, expanding the network whilst maintaining a c-split when compared to the GN2 default of 100 million jets improves the light-jet rejection but decreases the performance for c-jet rejection. This informs us that there is room for improvement when it comes to the GN2 network.

In Figure 5.7, introducing a c-split was explored and compared to samples with a default setup that has the same number of jets and 30 million jets. A trade-off between light- and c-jets was observed, and to further explore this trend, the jet rejection as a function of p_{T} at a fixed b-jet efficiency of 70% was plotted. In Figure 5.9, the distribution in the top panel depicts that in each bin, the GN2 c-split performs better for c-jet rejection when compared to the GN2 default sample. By introducing a c-split, the GN2 setup performs better at identifying and rejecting c-jets for $20 < p_{\mathrm{T}} < 250$ GeV. In the distribution in the bottom panel, at 70%, it is observed that GN2 default performs better for light-jet rejection when compared to the GN2 c-split sample for $20 < p_{\mathrm{T}} < 250$ GeV. As mentioned, this means that for GN2 set up by introducing information related to a specific flavour jet, it performs better at identifying and rejecting that flavour. This could, on the other hand, be a correction of mis-tagging but overall means that there is a need to work on improving this optimisation using the

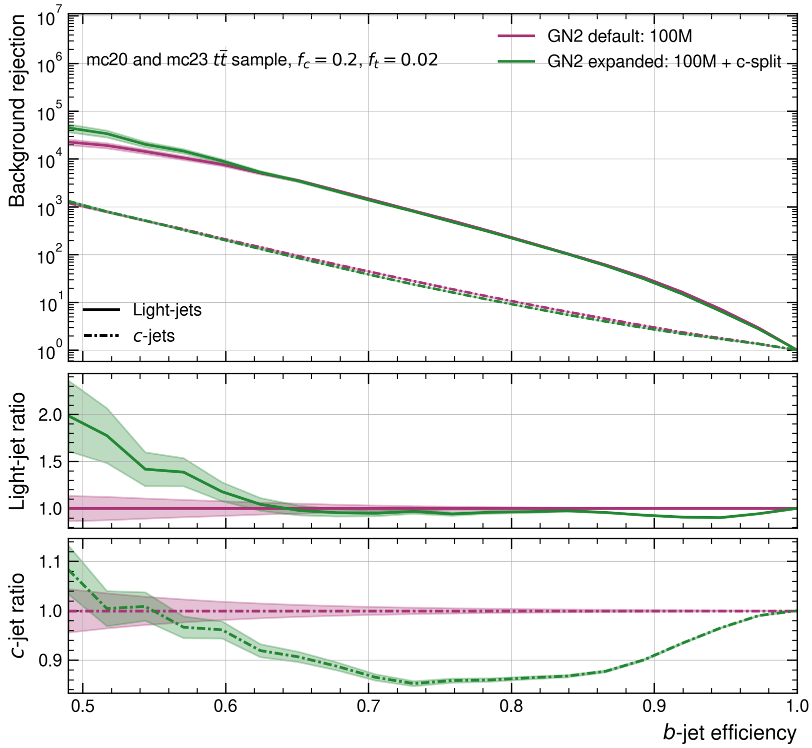

Figure 5.8: The c-jet, light-flavour jet , and b-jet rejections as a function of the b-jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_{\mathrm{T}} < 250$ GeV comparing two GN2 set-ups. One sample has 100 million jets and the other also has 100 million with c-split and increased number of layers and hidden layers in the network. The error bands are denoted by the shaded regions.

GN2 network.

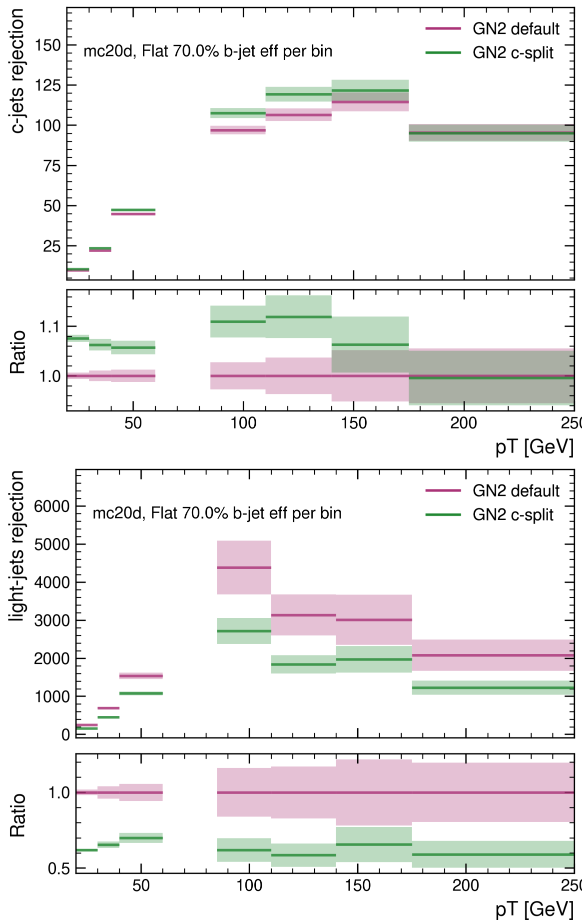

Figure 5.9: The c-jet rejection using $t\bar{t}$ sample (top) and light-jet rejection in the $t\bar{t}$ sample (bottom) as a function of jet p_{T} with a fixed b-jet efficiency of 70%. Binomial error bands are denoted by the shaded regions.

5.1.6 Conclusions

The idea around ML in HF tagging is designed to make predictions on the underlying physics. The algorithms are developed and trained to exploit specific properties and patterns, then validated and tested. Run 2 utilised a two-level tagger, where the outputs of low-level taggers are fed into high-level taggers as inputs. One of the disadvantages of the two-level tagger, as previously mentioned, is the dependence it has on low-level taggers and the difficulty of dealing with reconstructed charged particle tracks. This created a need for GN2 architecture. GN2 is based on GN1 architecture with an all-in-one transformer that has enhanced performance that can be used in different cases, such as trigger and upgrade. From the results obtained, increasing the number of jets in the training optimises GN2 models close to 1.2 times. This means more c-jet and light-jet rejections at 70% efficiency. Introducing a c-split in the sample uncovers an implicit trade-off between c-jets and light-jets, which could mean a correction of the mis-tagging or room for improving this optimisation using the GN2 network. Expanding the network whilst keeping the c-split reveals an unexpected trade-off, as the performance improves for light jets but decreases the performance for c-jets. Despite there being room for development for GN2, the network has displayed significant improvement in performance.

5.2 CERN Summer Programme: Performance Studies on Beauty and Charm Jets

5.2.1 Study Strategy

The aim of the project was to develop HF-jet tagging in ALICE. HF-jet tagging in the ALICE experiment involves identifying jets that originate from the hadronisation of HF quarks, specifically charm and bottom quarks, as previously mentioned. These HF jets are crucial for studying the properties of heavy quarks and their interactions with the quark-gluon plasma produced in high-energy heavy-ion collisions. These developments are a good starting point for Run 3 analyses and ALICE 3 projections. In ALICE, two methods are used in HF-jet tagging: secondary vertex reconstruction and particle identification. Secondary vertex reconstruction is usually done by reconstructing the HF hadron and its associated displaced secondary vertex. Since they have longer lifetimes, they will have displaced vertices as seen in Figure 5.10, making them easier to tag. Reconstructing the HF hadron is an ineffective method due to low reconstruction efficiency, and in the performance study, the plan was to tag the HF jets by identifying the displaced secondary vertex. Particle identification searches for specific particles that are known to decay from a B hadron, but this is also an ineffective strategy as a result of low branching ratios.

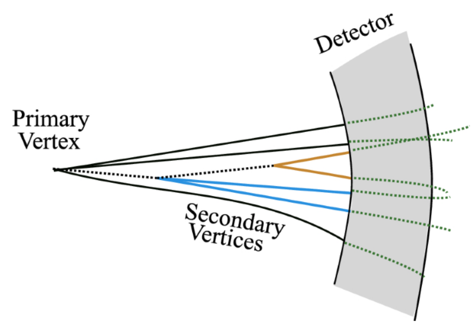

Figure 5.10: Diagram showing the primary vertex where the hard scattering of two protons or heavy-ions takes place and the secondary vertex is a result of the longer lifetime of HF hadrons. They travel a small distance before they decay into other particles

Using the tracks, the associated particle was found by working backwards. Once the particle was found, it was investigated to see if the mother was a b-hadron. The reason for doing this was that when the code would be run through the simulated data samples, the tagging performance could be assessed. The logic behind this approach is that b-quarks will fragment and hadronise into different variations of b-hadrons, so if the mother is either a B meson or B baryon, there is a chance the jet originated from a b-quark. Bottom quarks can come from the hard-scattered subprocess, subsequent subprocess, initial state radiation (ISR), beam remnants or gluon splitting. This meant that when computing the code, it had to identify the first mother but also check the PDG code [132] and the status code of the mother [133]. The PDG code would assist in isolating only the b-hadrons, and the status code would ensure that only events with b-quarks from the hard-scattered subprocess, subsequent subprocess, ISR, and beam remnants are chosen.

Two requirements for detector-level HF-jet tagging were employed. The first requirement was that once the associated particle was found using the track, the particles from hard-scattered subprocesses, subsequent subprocesses, ISR, and beam remnants were isolated using the relevant status codes: 23, 33, 43, and 63. The second requirement was from the isolated events; they were further required to have a specific PDG code which corresponds to b-mesons and b-baryons. The code was computed to run through a bottom-enriched simulated data for Run 3, and then a series of histograms were plotted to gain a better understanding of the kinematic properties of the mother particle and the resulting jet.

5.2.2 Simulated Samples Used

The MC event samples used were generated using `PYTHIA`. The detector-level MC event samples were obtained by passing it through `GEANT4` simulation that mimics the ALICE detector. The jets used in the study were reconstructed using tracks from charged particles and assembled using the anti-k_t algorithm with $R = 0.4$. Using the tracks, the particle associated with the track was found and then investigated to see if the mother is a b-hadron based on a PDG status code. The different status codes and their descriptions can be found in Ref. [133], and further explanation on execution will be discussed in Chapter 5.2. To analyse the results obtained, `ROOT` was used, and the FastJet package was used for jet clustering.

5.2.3 Results

A series of discriminating variables were plotted. In the first histogram, Figure 5.11, the entries in this plot are *b*-jets, and it shows the various status codes of their mothers. Close to half of the *b*-jets that were tagged originated from outgoing particles produced by final-state showers produced by parton branching (51). This is unexpected as bottom-enriched simulated data was used. The different status codes and their descriptions can be found in Ref. [133]. A cut was applied, and the jets were plotted. They were required to have a p_T of 10 GeV or more.

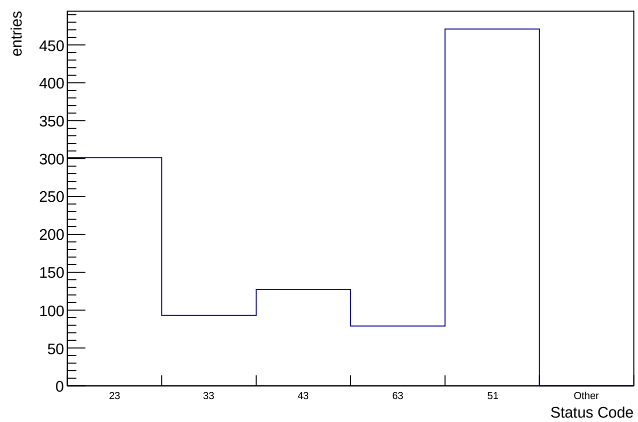

Figure 5.11: Histogram showing the various status codes of the mothers. The value 23 represents beauty jets whose mothers is an outgoing particle of the hardest subprocess, 43 are outgoing mothers produced by initial-state-showers through branching, 33 are outgoing mothers of subsequent subprocesses, 63 are outgoing mothers produced by beam-remnant treatment and lastly 51 are outgoing mothers produced by final-state showers produced by parton branching

In Figure 5.12, the histograms show the number of jets and *b*-jets in each collision. When comparing the plots, both peak at 0, this means that in each collision there are mostly no *b*-jets and jets.

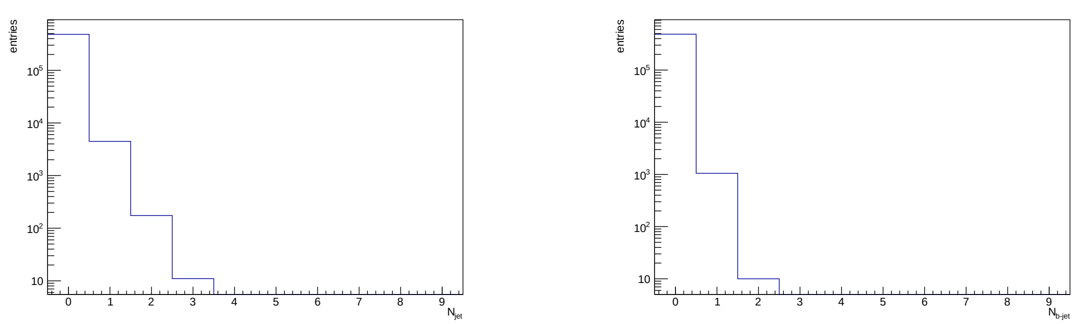

Figure 5.12: The histogram (a) on the left corresponds to the number of jets in each collision. The entries in this plot correspond to the number of collisions. Histogram (b) on the right displays the number of b-jets in each collision

In Figure 5.13, the histograms show the transverse momentum, p_T of jets and *b*-jets with a cut applied of 10 GeV. When comparing the plots, both exhibit a similar trend with a peak at 10 GeV followed by a steeply falling spectrum.

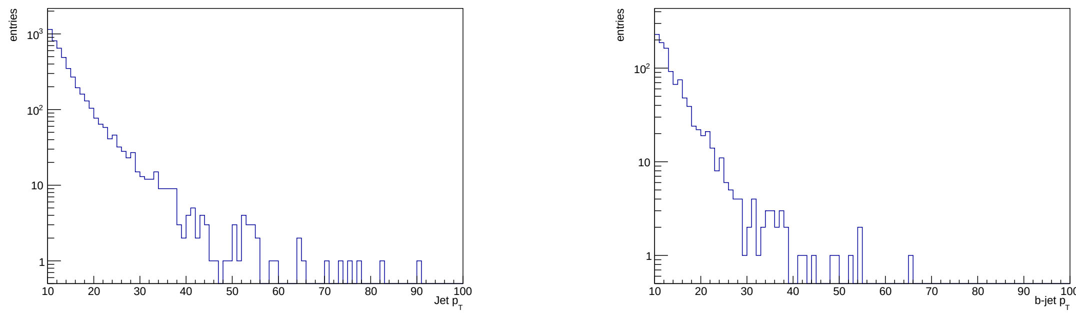

Figure 5.13: The histogram (a) on the left corresponds to the transverse momentum of all jets. Histogram (b) on the right also depicts the transverse momentum but of only b-jets.

Figure 5.14 displays the ratio of the p_{T} of the jet and the p_{T} of the mother. This plot allows for a deeper understanding of the underlying physics and analyses how much of the p_{T} was reconstructed. The plot peaks between 0 and 1, followed by a steeply falling spectrum. For jets reconstructed from tracks, it is expected to reconstruct about 2/3 of the initial parton momentum because most of the tracks are charged pions, and all pions are produced with the same probability.

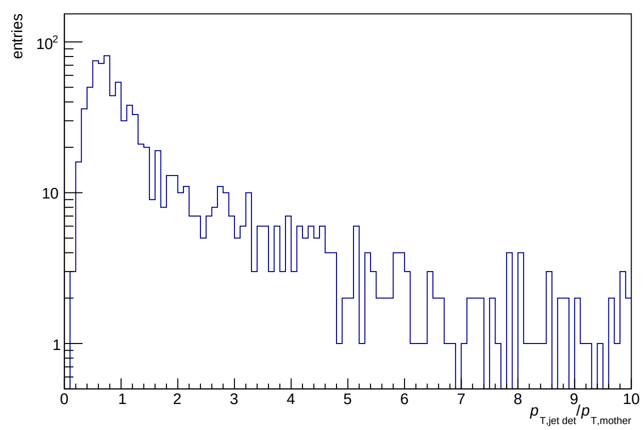

Figure 5.14: Histogram displaying the fraction between the p_{T} of the jet and the p_{T} of mother.

In Figure 5.15, plotted is the absolute value of the difference between the p_{T} of the jet and mother divided by the p_{T} of the jet. The distribution peaks at around 1, confirming that the jets plotted are charged jets.

In Figure 5.16, plotted is the absolute value of the difference between the η of the jet and the η of the mother. The distribution peaks at the value of 0.5, followed by a tail of values that extend to the value of 8.5. This was unexpected; the usual η values correspond to smaller values.

In Figure 5.17, plotted is the absolute value of the difference between the φ of the jet and the φ of the mother. The distribution peaks at zero followed by a tail of values.

5.2.4 Conclusion and Prospects

The aim of the project was to work on HF-jet tagging for ALICE for Run 3 analyses and ALICE 3 projections. The project was motivated by the difficulty of differentiating between a gluon and quark-initiated jet due to low branching fractions and low efficiency. In a broader aspect, HF-jet tagging

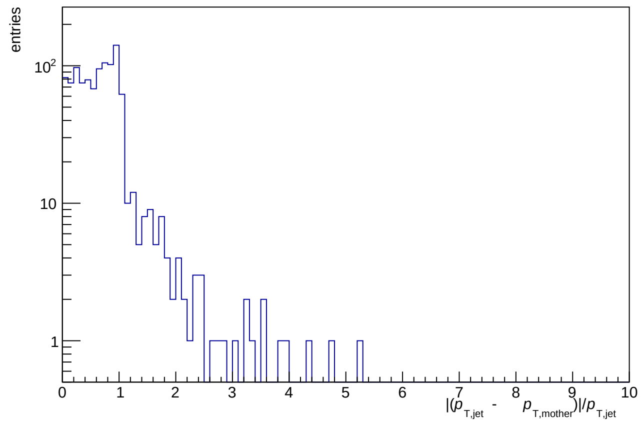

Figure 5.15: Histogram displaying the absolute value of the difference between the p_{T} of the jet and mother divided by the p_{T} of the jet.

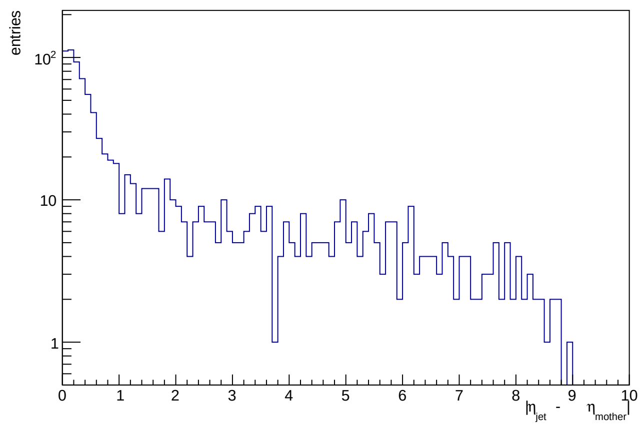

Figure 5.16: Histogram displaying the absolute value of the difference between the η of the jet and the η of the mother.

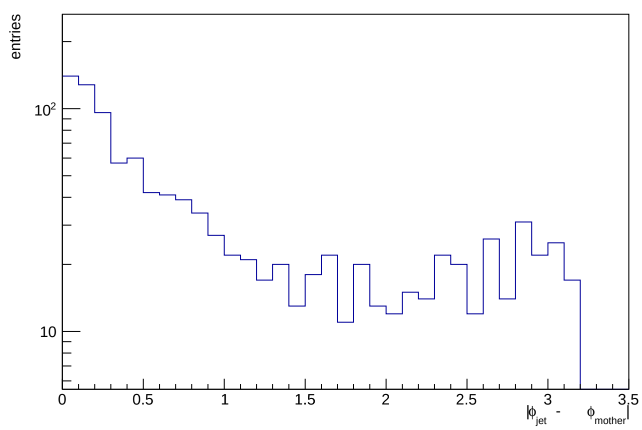

Figure 5.17: Histogram displaying the absolute value of the difference between the φ of the jet and the φ of the mother

will also assist in studying the production, showering and hadronisation mechanisms of heavy quarks in different environments, such as proton-proton or heavy-ion collisions. What has been achieved thus far is the first steps towards developing an MC criteria for detector-level HF-jet tagging in O^2 for b-jets. In the results that were obtained, there were a few unexpected trends related to the kinematic properties, such as the difference in η and why some of the mothers were outgoing particles produced by final-state showers produced by parton branching. This is unexpected based on the simulated data used. The next steps in this project would be developing a better understanding of the results obtained and modifying the HF-jet tagging task accordingly and extending the MC matching criteria to c-jets.

Chapter 6

Conclusion

The aim was to explore the different uses of b-jets at the LHC. As previously mentioned, detector final states have thousands of hadrons. Jets are not fundamental objects observed in final states but are constructed to deal with the large number of hadrons. In Chapter 2.3, the different jet clustering algorithms were presented. At the LHC, jets are cone-like structures due to the algorithm commonly used, known as anti-k_t with $R = 0.4$. For a jet to be classified as a b-jet at the particle level, there must be a b-hadron with a specific p_{T} value, whilst at the detector level it is more complicated and intricate because of the use of ML. What is common between the classification of b-jets at the particle and detector levels is that they exploit the properties of b-hadrons.

The answer to the question, What can be done with b-jets?" was presented in Chapters 3, 4, and 5. This displayed the versatility of b-jets. Their use is not limited to performance studies, measurements and searches. When it comes to searches, the b-jets were particularly useful for a BSM search. The exploration of bottom-quark-philic SVJs was a phenomenology study for a collider signature for DM. The SVJ-b signature was a continuation from the search for SVJ. The reason for reproducing the signature with b-jets was to make the SVJ signature a much less ambiguous collider signature, which assisted in reducing the background significantly. A new jet clustering algorithm called variable-R jets was used, and a set of requirements for the search strategy was recommended.

For the measurement, a b-jet cross-section only using events with high E_T^{miss} was presented after noting that there was no b-jet cross-section result using Run 2 data from ATLAS. A set of measured observables were defined and corrected for detector effects using the RooUnfold framework so they could be directly compared with theoretical distributions. The unfolding was done using the leading SM contribution, $t\bar{t}$ (non-had), generated from two event generators, namely, `SHERPA` and `PYTHIA`. From the results, it was observed that the `PYTHIA` event generator was in better agreement with the theoretical prediction.

What enables b-jets to be used in measurements and searches is HF tagging algorithms that are developed by the different experiments at the LHC. What these HF algorithms have in common in the different experiments is that they exploit different properties of b-hadrons; for example, they look for secondary vertices or signs of hard fragmentation. In a performance study done in ALICE, the aim of the project was to work on HF-jet tagging for ALICE for Run 3 analyses and ALICE 3 projections. An MC matching criteria for detector level HF-jet tagging in O^2 for b-jets was developed. The results obtained are to assist in studying the production, showering and hadronisation mechanisms of heavy quarks in different environments, such as proton-proton or heavy-ion collisions.

The HF tagging performance study in ATLAS was to optimise GN2 architecture, which is an ML algorithm that tags b-jets. The optimisation strategy uncovered that increasing the number of jets used in the training and increasing the GN2 network by adding an additional hidden layer improves the performance of the GN2 network by up to 1.2 times. An interesting phenomenon was discovered by introducing a c-jet class split into the GN2 training. The samples used uncovered an implicit trade-off between c-jets and light-jets. The network has displayed significant improvement in performance and is on track to be the recommended ATLAS b-tagging algorithm in the future.

Bibliography

[1] M. K. Gaillard, P. D. Grannis, F. J. Sciulli, The Standard Model of Particle Physics, Reviews of Modern Physics 71 (2) (1999) S96. doi:{10.1103/revmodphys.71.s96}.
URL http://dx.doi.org/10.1103/RevModPhys.71.S96

[2] W. N. Cottingham, D. A. Greenwood, An Introduction to the Standard Model of Particle Physics, Cambridge University Press, 2007. doi:10.1017/9781009401685.

[3] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Physics Letters B 716 (1) (2012) 30–61. doi:10.1016/j.physletb.2012.08.021.
URL http://dx.doi.org/10.1016/j.physletb.2012.08.021

[4] ATLAS Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B 716 (1) (2012) 1–29. doi:10.1016/j.physletb.2012.08.020.
URL http://dx.doi.org/10.1016/j.physletb.2012.08.020

[5] M. Robinson, Symmetry and The Standard Model: Mathematics and Particle Physics , Springer, 2011. doi:10.1007/978-1-4419-8267-4.

[6] Quantum Diaries, The Standard Model: A Beautiful But Flawed Theory, Accessed: 2022-11-04.
URL https://www.quantumdiaries.org/2014/03/14/the-standard-model-a-beautiful-but-flawed-theory/

[7] D. Kar, Experimental Particle Physics, IOP Publishing Limited, 2019. doi:10.1088/2053-2563/ab1be6.

[8] C. Gómez, R. Jimenez, The Quantum Cosmological Tilt and the Origin of Dark Matter, Journal of Cosmology and Astroparticle Physics (10) (2021) 055. arXiv:2012.07883, doi:10.1088/1475-7516/2021/10/055.

[9] J. C. Walter, G. T. Barkema, An introduction to Monte Carlo methods, Physica A: Statistical Mechanics and its Applications 418 (2015) 78–87. doi:10.1016/j.physa.2014.06.014.
URL http://dx.doi.org/10.1016/j.physa.2014.06.014

[10] J. M. Hammersley, Monte Carlo Methods for Solving Multivariable Problems, Vol. 86, The New York Academy of Sciences, 1960, pp. 844–874.
URL https://doi.org/10.1111/j.1749-6632.1960.tb42846.x

[11] T. Sjöstrand, S. Mrenna, P. Z. Skands, A Brief Introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852–867. arXiv:0710.3820, doi:10.1016/j.cpc.2008.01.036.

[12] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, P. Z. Skands, An Introduction to PYTHIA 8.2, Comput. Phys. Commun. 191 (2015) 159–177. arXiv:1410.3012, doi:10.1016/j.cpc.2015.01.024.

[13] T. Gleisberg, et al, SHERPA 1. alpha: A proof-of-concept version, Journal of High Energy Physics 2004 (02) (2004) 056–056. doi:10.1088/1126-6708/2004/02/056.
URL http://dx.doi.org/10.1088/1126-6708/2004/02/056

[14] G. Corcella, et al., HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes), JHEP 01 (2001) 010. arXiv:hep-ph/0011363, doi:10.1088/1126-6708/2001/01/010.

[15] G. Marchesini, et al., HERWIG version 5.9 (1996). arXiv:hep-ph/9607393.
URL https://arxiv.org/abs/hep-ph/9607393

[16] P. Nason, A new method for combining NLO QCD with shower Monte Carlo algorithms, Journal of High Energy Physics 2004 (11) (2004) 040. doi:10.1088/1126-6708/2004/11/040.
URL http://dx.doi.org/10.1088/1126-6708/2004/11/040

[17] S. Frixione, P. Nason, C. Oleari, Matching NLO QCD computations with parton shower simulations: the POWHEG method, Journal of High Energy Physics 2007 (11) (2007) 070. doi:10.1088/1126-6708/2007/11/070.
URL http://dx.doi.org/10.1088/1126-6708/2007/11/070

[18] S. Alioli, P. Nason, C. Oleari, E. Re, A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX, Journal of High Energy Physics 2010 (6) (2010) 1–58. doi:10.1007/jhep06(2010)043.
URL http://dx.doi.org/10.1007/JHEP06(2010)043

[19] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP 07 (2014) 079. arXiv:1405.0301, doi:10.1007/JHEP07(2014)079.

[20] J. C. Collins, D. E. Soper, G. Sterman, Factorization of Hard Processes in QCD (2004). arXiv:hep-ph/0409313.
URL https://arxiv.org/abs/hep-ph/0409313

[21] A. D. Martin, W. J. Stirling, R. S. Thorne, G. Watt, Parton distributions for the LHC, The European Physical Journal C (2009). doi:10.1140/epjc/s10052-009-1072-5.
URL https://doi.org/10.1140/epjc/s10052-009-1072-5

[22] I. W. Stewart, F. J. Tackmann, W. J. Waalewijn, Factorization at the LHC: From parton distribution functions to initial state jets, Physical Review D 81 (9) (May 2010). doi:10.1103/physrevd.81.094035.
URL http://dx.doi.org/10.1103/PhysRevD.81.094035

[23] R. D. Ball, et al., A first unbiased global NLO determination of parton distributions and their uncertainties, Nuclear Physics B 838 (1–2) (2010) 136–206. doi:10.1016/j.nuclphysb.2010.05.008.
URL http://dx.doi.org/10.1016/j.nuclphysb.2010.05.008

[24] M. R. Whalley, D. Bourilkov, R. C. Group, The Les Houches Accord PDFs (LHAPDF) and Lhaglue (2005). arXiv:hep-ph/0508110.
URL https://arxiv.org/abs/hep-ph/0508110

[25] A. Buckley, et al., LHAPDF6: Parton Density Access in the LHC Precision Era, The European Physical Journal C 75 (3) (2015). doi:10.1140/epjc/s10052-015-3318-8.
URL http://dx.doi.org/10.1140/epjc/s10052-015-3318-8

[26] J. M. Campbell, R. K. Ellis, C. Williams, Vector Boson Pair Production at the LHC, Journal of High Energy Physics 2011 (7) (Jul. 2011). doi:10.1007/jhep07(2011)018.
URL http://dx.doi.org/10.1007/JHEP07(2011)018

[27] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, B. Mistlberger, Higgs Boson Gluon-fusion Production Beyond Threshold in N3LO QCD (2014). arXiv:1411.3584.
URL https://arxiv.org/abs/1411.3584

[28] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Höche, H. Hoeth, F. Krauss, L. Lönnblad, E. Nurse, P. Richardson, S. Schumann, M. H. Seymour, T. Sjöstrand, P. Skands, B. Webber, General-purpose event generators for LHC physics, Physics Reports 504 (5) (2011)

145–233. doi:10.1016/j.physrep.2011.03.005.
URL http://dx.doi.org/10.1016/j.physrep.2011.03.005

[29] Z. Nagy, D. E. Soper, What is a parton shower?, Phys. Rev. D 98 (2018) 014034. doi:10.1103/PhysRevD.98.014034.
URL https://link.aps.org/doi/10.1103/PhysRevD.98.014034

[30] S. Höche, Introduction to parton-shower event generators (2015). arXiv:1411.4085.
URL https://arxiv.org/abs/1411.4085

[31] S. Hoeche, F. Krauss, N. Lavesson, L. Lonnblad, M. Mangano, A. Schaelicke, S. Schumann, Matching Parton Showers and Matrix Elements (2006). arXiv:hep-ph/0602031.
URL https://arxiv.org/abs/hep-ph/0602031

[32] M. L. Mangano, M. Moretti, F. Piccinini, M. Treccani, Matching matrix elements and shower evolution for top-quark production in hadronic collisions, JHEP 01 (2007) 013. arXiv:hep-ph/0611129, doi:10.1088/1126-6708/2007/01/013.

[33] B. Andersson, G. Gustafson, G. Ingelman, T. Sjostrand, Parton Fragmentation and String Dynamics, Phys. Rept. 97 (1983) 31–145. doi:10.1016/0370-1573(83)90080-7.

[34] B. W. G. Marchesini, Simulation of QCD jets including soft gluon interference, Nuclear Physics B 238 (1984) 1–29.
URL https://doi.org/10.1016/0550-3213(84)90463-2

[35] B. Webber, A QCD model for jet fragmentation including soft gluon interference, Nuclear Physics B 238 (1984) 492–528.
URL https://doi.org/10.1016/0550-3213(84)90333-X

[36] B. Andersson, G. Gustafson, G. Ingelman, T. Sjöstrand, Parton fragmentation and string dynamics, Physics Reports 97 (2) (1983) 31–145. doi:10.1016/0370-1573(83)90080-7.
URL https://doi.org/10.1016/0370-1573(83)90080-7

[37] B. Andersson, The Lund Model, Vol. 7, Cambridge University Press, 1998. doi:10.1017/9781009401296.
URL https://doi.org/10.1017/9781009401296

[38] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, J. E. von Seggern, Systematic Event Generator Tuning for the LHC, The European Physical Journal C 65 (1–2) (2009) 331–357. doi:10.1140/epjc/s10052-009-1196-7.
URL http://dx.doi.org/10.1140/epjc/s10052-009-1196-7

[39] ATLAS Collaboration, ATLAS Pythia 8 tunes to 7 TeV data, ATL-PHYS-PUB-2014-021 (2014).
URL https://cds.cern.ch/record/1966419

[40] ATLAS Collaboration, The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model, ATL-PHYS-PUB-2016-017 (2016).
URL https://cds.cern.ch/record/2206965

[41] GEANT4 Collaboration, S. Agostinelli, et al., GEANT4 - a simulation toolkit, Nucl. Instrum. Meth. A 506 (25) (2003) 0.
URL https://doi.org/10.1016/S0168-9002(03)01368-8

[42] J. B. H. Matt Dobbs, The HepMC C++ Monte Carlo event record for High Energy , Computer Physics Communications 134 (2001) 41–46.
URL https://doi.org/10.1016/S0010-4655(00)00189-2

[43] J. Alwall, et al., A standard format for Les Houches Event Files, Computer Physics Communications 176 (2007) 300–304.
URL https://doi.org/10.1016/j.cpc.2006.11.010

[44] ATLAS Collaboration, The ATLAS experiment at the CERN Large Hadron Collider, Journal of Instrumentation (2008). doi:10.1088/1748-0221/3/08/S08003.
URL https://doi.org/10.1088/1748-0221/3/08/S08004

[45] ATLAS Collaboration, Differential cross-sections for events with missing transverse momentum and jets measured with the ATLAS detector in 13 TeV proton-proton collisions, Journal of High Energy Physics (2024). arXiv:2403.02793, doi:10.1007/jhep08(2024)223.
URL http://dx.doi.org/10.1007/JHEP08(2024)223

[46] G. Avoni, et al., The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS, JINST 13 (07) (2018) P07017. doi:10.1088/1748-0221/13/07/P07017.
URL https://cds.cern.ch/record/2633501

[47] ALICE Collaboration, The ALICE experiment at the CERN LHC, Journal of Instrumentation 3 (08) (2008) S08002. doi:10.1088/1748-0221/3/08/S08002.
URL https://doi.org/10.1088/1748-0221/3/08/S08002

[48] ATLAS Collaboration, Performance of the ATLAS trigger system in 2015, Eur. Phys. J. C 77 (2017) 317. arXiv:1611.09661, doi:10.1140/epjc/s10052-017-4852-3.

[49] ATLAS Collaboration, Jet reconstruction and performance using particle flow with the ATLAS Detector, The European Physical Journal C 77 (2017) 1–47. doi:10.1140/epjc/s10052-017-5031-2.
URL http://dx.doi.org/10.1140/epjc/s10052-017-5031-2

[50] M. Cacciari, G. P. Salam, G. Soyez, The anti-k_t jet clustering algorithm, JHEP 04 (2008) 063. arXiv:0802.1189, doi:10.1088/1126-6708/2008/04/063.

[51] M. Cacciari, G. P. Salam, G. Soyez, FastJet User Manual, Eur. Phys. J. C72 (2012) 1896. arXiv:1111.6097, doi:10.1140/epjc/s10052-012-1896-2.

[52] J. Collins, T. Rogers, On the Definition of Fragmentation Functions and the Violation of Sum Rules (2024). arXiv:2309.03346.
URL https://arxiv.org/abs/2309.03346

[53] ATLAS Collaboration, ATLAS b-jet Identification Performance and Efficiency Measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV, The European Physical Journal C 79 (11) (Jan 2020). doi:10.1140/epjc/s10052-019-7450-8.
URL http://dx.doi.org/10.1140/epjc/s10052-019-7450-8

[54] ATLAS Collaboration, Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV, The European Physical Journal C (2018). arXiv:1802.08168, doi:10.1140/epjc/s10052-018-6288-9.
URL http://dx.doi.org/10.1140/epjc/s10052-018-6288-9

[55] T. Cohen, M. Lisanti, H. K. Lou, S. Mishra-Sharma, LHC Searches for Dark Sector Showers, JHEP 11 (2017) 196. arXiv:1707.05326, doi:10.1007/JHEP11(2017)196.

[56] S. Profumo, An Introduction to Particle Dark Matter, World Scientific Publishing Company, 2017. arXiv:1910.05610.
URL https://arxiv.org/abs/1910.05610

[57] S. Dodelson, The Real Problem with MOND, International Journal of Modern Physics D 20 (14) (2011) 2749–2753. doi:10.1142/s0218271811020561.
URL http://dx.doi.org/10.1142/S0218271811020561

[58] L. Wyrzykowski, et al., The OGLE view of microlensing towards the Magellanic Clouds–IV. OGLE-III SMC data and final conclusions on MACHOs, Monthly Notices of the Royal Astronomical Society 416 (4) (2011) 2949–2961. doi:10.1111/j.1365-2966.2011.19243.x.
URL http://dx.doi.org/10.1111/j.1365-2966.2011.19243.x

[59] P. Tisserand, et al., Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds, Astronomy & Astrophysics 469 (2) (2007) 387–404. doi:10.1051/00

04-6361:20066017.
URL http://dx.doi.org/10.1051/0004-6361:20066017

[60] B. W. Lee, S. Weinberg, Cosmological Lower Bound on Heavy-neutrino Masses, Physical Review Letters 39 (4) (1977) 165. doi:10.1103/PhysRevLett.39.165.
URL https://link.aps.org/doi/10.1103/PhysRevLett.39.165

[61] F. Reines, The Neutrino: From Poltergeist to Particle, Reviews of Modern Physics 68 (2) (1996) 317. doi:10.1103/RevModPhys.68.317.
URL https://link.aps.org/doi/10.1103/RevModPhys.68.317

[62] J. R. Primack, D. Seckel, B. Sadoulet, Detection of Cosmic Dark Matter, Annual Review of Nuclear and Particle Science 38 (1) (1988) 751–807. doi:10.1146/annurev.ns.38.120188.003535.

[63] G. Barello, S. Chang, C. A. Newby, A Model Independent Approach to Inelastic Dark Matter scattering, Physical Review D 90 (9) (2014) 094027. doi:10.1103/physrevd.90.094027.
URL http://dx.doi.org/10.1103/PhysRevD.90.094027

[64] M. Cirelli, A. Strumia, M. Tamburini, Cosmology and Astrophysics of Minimal Dark Matter, Nuclear Physics B 787 (1-2) (2007) 152–175. doi:10.1016/j.nuclphysb.2007.07.023.
URL http://dx.doi.org/10.1016/j.nuclphysb.2007.07.023

[65] Y. A. Gol'fand, E. P. Likhtman, Extension of the algebra of Poincaré group generators and violation of P invariance, in: Supergravities in diverse dimensions. Volume 1, World Scientific, 1989. doi:10.1142/9789814542340_0001.
URL https://doi.org/10.1142/9789814542340_0001

[66] A. Salam, J. Strathdee, Super-symmetry and non-Abelian Gauges, Physics Letters B 51 (4) (1974) 353–355. doi:10.1016/0370-2693(74)90226-3.
URL https://doi.org/10.1142/9789812795915_0050

[67] J. Wess, B. Zumino, Supergauge Invariant Extension of Quantum Electrodynamics, Nuclear Physics B 78 (1) (1974) 1–13. doi:10.1016/0550-3213(74)90112-6.
URL https://doi.org/10.1016/0550-3213(74)90112-6

[68] Y. Mambrini, S. Profumo, F. S. Queiroz, Dark Matter and Global Symmetries, Physics Letters B 760 (2016) 807–815. doi:10.1016/j.physletb.2016.07.076.
URL http://dx.doi.org/10.1016/j.physletb.2016.07.076

[69] S. Colafrancesco, S. Profumo, P. Ullio, Multi-frequency Analysis of Neutralino Dark Matter Annihilations in the Coma cluster, Astronomy & Astrophysics 455 (1) (2006) 21–43. doi:10.1051/0004-6361:20053887.
URL http://dx.doi.org/10.1051/0004-6361:20053887

[70] A. Boveia, C. Doglioni, Dark Matter Searches at Colliders, Annual Review of Nuclear and Particle Science 68 (2018) 429–459. doi:10.1146/annurev-nucl-101917-021008.
URL http://dx.doi.org/10.1146/annurev-nucl-101917-021008

[71] T. Cohen, M. Lisanti, H. K. Lou, Semivisible Jets: Dark Matter Undercover at the LHC, Phys. Rev. Lett. 115 (17) (2015) 171804. arXiv:1503.00009, doi:10.1103/PhysRevLett.115.171804.

[72] E. Bernreuther, F. Kahlhoefer, M. Krämer, P. Tunney, Strongly interacting dark sectors in the early Universe and at the LHC through a simplified portal, JHEP 01 (2020) 162. arXiv:1907.04346, doi:10.1007/JHEP01(2020)162.

[73] G. Albouy, et al., Theory, phenomenology, and experimental avenues for dark showers: a Snowmass 2021 report (3 2022). arXiv:2203.09503.

[74] ATLAS Collaboration, Search for non-resonant production of semi-visible jets using Run 2 data in ATLAS, Physics Letters B (5 2023). arXiv:2305.18037, doi:10.1016/j.physletb.2023.138324.
URL http://dx.doi.org/10.1016/j.physletb.2023.138324

[75] CMS Collaboration, Search for resonant production of strongly coupled dark matter in proton–proton collisions at 13 TeV (2021). arXiv:2112.11125.

[76] Analysis Software Group, Introduction to MET, Accessed: 2023-06-08 (2023).
URL https://atlassoftwaredocs.web.cern.ch/AnalysisSWTutorial/obj_met_intro/

[77] M. J. Strassler, K. M. Zurek, Echoes of a hidden valley at hadron colliders, Phys. Lett. B 651 (2007) 374–379. arXiv:hep-ph/0604261, doi:10.1016/j.physletb.2007.06.055.

[78] M. J. Strassler, On the Phenomenology of Hidden Valleys with Heavy Flavor, (6 2008). arXiv:0806.2385.

[79] S. Renner, P. Schwaller, A flavoured dark sector, JHEP 08 (2018) 052. arXiv:1803.08080, doi:10.1007/JHEP08(2018)052.

[80] C. A. Manzari, S. Profumo, A Flavour Inspired Model for Dark Matter (6 2022). arXiv:2206.06768.

[81] ATLAS Collaboration, Configuration and Performance of the ATLAS b-jet Triggers in Run 2, The European Physical Journal C 81 (12) (2021) 1087. doi:10.1140/epjc/s10052-021-09775-5.
URL http://dx.doi.org/10.1140/epjc/s10052-021-09775-5

[82] T. Lou, S. Mishra-Sharma, Dark Matter Simp Model, https://github.com/smsharma/SemivisibleJets, Accessed: 2020-11-10.

[83] NNPDF Collaboration, Parton distributions for the LHC Run II, Journal of High Energy Physics 2015 (4) (2015) 1–148. doi:10.1007/jhep04(2015)040.
URL http://dx.doi.org/10.1007/JHEP04(2015)040

[84] L. Carloni, T. Sjöstrand, Visible Effects of Invisible Hidden Valley Radiation, JHEP 09 (2010) 105. arXiv:1006.2911, doi:10.1007/JHEP09(2010)105.

[85] L. Carloni, J. Rathsman, T. Sjöstrand, Discerning Secluded Sector gauge structures, JHEP 04 (2011) 091. arXiv:1102.3795, doi:10.1007/JHEP04(2011)091.

[86] T. Cohen, M. Lisanti, H. K. Lou, S. Mishra-Sharma, LHC searches for dark sector showers, Journal of High Energy Physics 2017 (11) (2017) 1–32. arXiv:1707.05326, doi:10.1007/JHEP11(2017)196.

[87] C. Bierlich, et al., Robust Independent Validation of Experiment and Theory: Rivet version 3, SciPost Phys. 8 (2020) 026. arXiv:1912.05451, doi:10.21468/SciPostPhys.8.2.026.

[88] ATLAS Collaboration, Search for Supersymmetry in Final States with Missing Transverse Momentum and Three or more b-jets in 139 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, The European Physical Journal C (2022). doi:10.1140/epjc/s10052-023-11543-6.
URL http://dx.doi.org/10.1140/epjc/s10052-023-11543-6

[89] ATLAS Collaboration, Search for dark matter produced in association with a Standard Model Higgs boson decaying into b-quarks using the full Run 2 dataset from the ATLAS detector, Journal of High Energy Physics (2021). doi:10.1007/jhep11(2021)209.
URL http://dx.doi.org/10.1007/JHEP11(2021)209

[90] E. Maguire, L. Heinrich, G. Watt, HEPData: a repository for high energy physics data, Journal of Physics: Conference Series 898 (2017) 102006. doi:10.1088/1742-6596/898/10/102006.
URL http://dx.doi.org/10.1088/1742-6596/898/10/102006

[91] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, J. P. Silva, Theory and phenomenology of two-Higgs-doublet models, Physics Reports 516 (1–2) (2012) 1–102. doi:10.1016/j.physrep.2012.02.002.
URL http://dx.doi.org/10.1016/j.physrep.2012.02.002

[92] A. Berlin, T. Lin, L.-T. Wang, Mono-Higgs detection of dark matter at the LHC, Journal of High Energy Physics 2014 (6) (Jun. 2014). doi:10.1007/jhep06(2014)078.
URL http://dx.doi.org/10.1007/JHEP06(2014)078

[93] LHC Dark Matter Working Group, LHC Dark Matter Working Group: Next-generation spin-0 dark matter models (2018). arXiv:1810.09420.
URL https://arxiv.org/abs/1810.09420

[94] M. Bauer, U. Haisch, F. Kahlhoefer, Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators, Journal of High Energy Physics 2017 (5) (May 2017). doi:10.1007/jhep05(2017)138.
URL http://dx.doi.org/10.1007/JHEP05(2017)138

[95] ATLAS Collaboration, Search for dark matter produced in association with a Standard Model Higgs boson decaying into b-quarks using the full Run 2 dataset from the ATLAS detector, Journal of High Energy Physics (2021). doi:10.1007/jhep11(2021)209.
URL http://dx.doi.org/10.1007/JHEP11(2021)209

[96] D. Kar, S. Sinha, Exploring jet substructure in semi-visible jets, SciPost Phys. 10 (2021) 084. doi:10.21468/SciPostPhys.10.4.084.
URL https://scipost.org/10.21468/SciPostPhys.10.4.084

[97] ATLAS Collaboration, Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions, The European Physical Journal C 81 (4) (Apr. 2021). doi:10.1140/epjc/s10052-021-09054-3.
URL http://dx.doi.org/10.1140/epjc/s10052-021-09054-3

[98] D. Krohn, J. Thaler, L.-T. Wang, Jet trimming, Journal of High Energy Physics 2010 (2) (Feb. 2010). doi:10.1007/jhep02(2010)084.
URL http://dx.doi.org/10.1007/JHEP02(2010)084

[99] B. Nachman, P. Nef, A. Schwartzman, M. Swiatlowski, C. Wanotayaroj, Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC, JHEP 02 (2015) 075. arXiv:1407.2922, doi:10.1007/JHEP02(2015)075.

[100] B. Mukhopadhyaya, T. Samui, R. K. Singh, Dynamic radius jet clustering algorithm, Journal of High Energy Physics 2023 (4) (Apr. 2023). doi:10.1007/jhep04(2023)019.
URL http://dx.doi.org/10.1007/JHEP04(2023)019

[101] D. Krohn, J. Thaler, L.-T. Wang, Jets with variable R, Journal of High Energy Physics 2009 (06) (2009) 059–059. doi:10.1088/1126-6708/2009/06/059.
URL http://dx.doi.org/10.1088/1126-6708/2009/06/059

[102] A. Chakraborty, S. Dasmahapatra, H. A. Day-Hall, B. Ford, S. Jain, S. Moretti, Fat b-jet analyses using old and new clustering algorithms in new Higgs boson searches at the LHC, The European Physical Journal C 83 (4) (Apr. 2023). doi:10.1140/epjc/s10052-023-11537-4.
URL http://dx.doi.org/10.1140/epjc/s10052-023-11537-4

[103] D. Kar, W. Nzuza, S. Sinha, 2B or not 2B, a study of bottom-quark-philic semi-visible jets (2024). arXiv:2207.01885.
URL https://arxiv.org/abs/2207.01885

[104] ATLAS Collaboration, Differential cross-sections for events with missing transverse momentum and jets measured with the atlas detector in 13 tev proton-proton collisions, Journal of High Energy Physics 2024 (8) (Aug. 2024). doi:10.1007/jhep08(2024)223.
URL http://dx.doi.org/10.1007/JHEP08(2024)223

[105] NNPDF Collaboration, Parton distributions for the LHC Run II, Journal of High Energy Physics 2015 (4) (2015) 1–148. doi:10.1007/jhep04(2015)040.
URL http://dx.doi.org/10.1007/JHEP04(2015)040

[106] D. J. Lange, The EvtGen particle decay simulation package, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 462 (1-2) (2001) 152–155.
URL https://doi.org/10.1016/S0168-9002(01)00089-4

[107] E. Bothmann, et al., Event generation with Sherpa 2.2, SciPost Physics 7 (3) (2019) 034. doi:10.21468/scipostphys.7.3.034.
URL http://dx.doi.org/10.21468/SciPostPhys.7.3.034

[108] ATLAS Collaboration, The ATLAS simulation infrastructure, The European Physical Journal C 70 (3) (2010) 823–874. doi:10.1140/epjc/s10052-010-1429-9.
URL http://dx.doi.org/10.1140/epjc/s10052-010-1429-9

[109] ATLAS Collaboration, Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1, The European Physical Journal C 77 (7) (2017) 1–73. doi:10.1140/epjc/s10052-017-5004-5.
URL http://dx.doi.org/10.1140/epjc/s10052-017-5004-5

[110] ATLAS collaboration, Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton-proton run, Journal of Instrumentation (2013). doi:10.1088/1748-0221/8/07/p07004.
URL http://dx.doi.org/10.1088/1748-0221/8/07/P07004

[111] ATLAS Collaboration, Performance of the missing transverse momentum triggers for the ATLAS detector during Run-2 data taking, Journal of High Energy Physics 2020 (8) (Aug. 2020). doi:10.1007/jhep08(2020)080.
URL http://dx.doi.org/10.1007/JHEP08(2020)080

[112] L. Brenner, P. Verschuuren, R. Balasubramanian, C. Burgard, V. Croft, G. Cowan, W. Verkerke, Comparison of unfolding methods using RooFitUnfold, World Scientific (2020). arXiv:1910.14654.
URL https://arxiv.org/abs/1910.14654

[113] T. Adye, Unfolding algorithms and tests using RooUnfold (2011). arXiv:1105.1160.
URL https://arxiv.org/abs/1105.1160

[114] G. D'Agostini, Improved iterative Bayesian unfolding (2010). arXiv:1010.0632.
URL https://arxiv.org/abs/1010.0632

[115] ATLAS Collaboration, Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{13}$ TeV with the ATLAS detector, Physical Review D 96 (7) (Oct. 2017). doi:10.1103/physrevd.96.072002.
URL http://dx.doi.org/10.1103/PhysRevD.96.072002

[116] ATLAS Collaboration, Study of the material of the ATLAS inner detector for Run 2 of the LHC, Journal of Instrumentation 12 (12) (2017). doi:10.1088/1748-0221/12/12/p12009.
URL http://dx.doi.org/10.1088/1748-0221/12/12/P12009

[117] ATLAS Collaboration, Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2, The European Physical Journal C 77 (10) (2017). doi:10.1140/epjc/s10052-017-5225-7.
URL http://dx.doi.org/10.1140/epjc/s10052-017-5225-7

[118] ATLAS Collaboration, Measurement of track reconstruction inefficiencies in the core of jets via pixel dE/dx with the ATLAS experiment using $\sqrt{s} = 13$ TeV pp collision data, Tech. rep., CERN, Geneva (2016).
URL https://cds.cern.ch/record/2140460

[119] ATLAS Collaboration, Modelling of Track Reconstruction Inside Jets with the 2016 ATLAS $\sqrt{s} = 13$ TeV pp dataset, Tech. rep., CERN, Geneva, all figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-016 (2017).
URL https://cds.cern.ch/record/2275639

[120] ATLAS Collaboration, ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV, The European Physical Journal C 79 (11) (Nov. 2019). doi:10.1140/epjc/s10052-019-7450-8.
URL http://dx.doi.org/10.1140/epjc/s10052-019-7450-8

[121] ATLAS Collaboration, Measurement of b-tagging Efficiency of c-jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector, Tech. rep., CERN (2018).
URL https://cds.cern.ch/record/2306649/files/ATLAS-CONF-2018-001.pdf

[122] ATLAS COllaborations, Performance of the ATLAS muon triggers in Run 2, Journal of Instrumentation 15 (09) (Sep. 2020). doi:10.1088/1748-0221/15/09/p09015.
URL http://dx.doi.org/10.1088/1748-0221/15/09/p09015

[123] ATLAS Collaboration, Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC, The European Physical Journal C 83 (10) (2023). doi:10.1140/epjc/s10052-023-11747-w.
URL http://dx.doi.org/10.1140/epjc/s10052-023-11747-w

[124] C.-Y. Wong, Introduction to high-energy heavy-ion collisions, World Scientific, 1994.
URL https://doi.org/10.1142/1128

[125] ATLAS Collaboration, Identifying Jet Flavour with Transformers in LHC Run-3 ATLAS p-p Collision Data, ATLAS-PUB-NOTE (June 2023).
URL https://cds.cern.ch/record/2811135/files/ATL-PHYS-PUB-2022-027.pdf

[126] N. Kumari, Flavour Tagging with Graph Neural Network with the ATLAS Detector, in: BOOST 2024, 2024, BOOST 2024 Genova, Italy 29th July 2024.
URL https://agenda.infn.it/event/37093/contributions/234315/attachments/124111/182400/BOOST24_ftag_atlas_kumari.pdf

[127] S. V. Stroud, Flavour Tagging with Graph Neural Networks at ATLAS, CERN Data Science Seminar (2023).
URL https://indico.cern.ch/event/1232499/attachments/2602341/4494127/2023-03-01_GN1_Seminar.pdf

[128] N. Kumari, K. Peters, Extended Flavour Labelling Scheme for GN2 b-tagger, in: FTAG algorithms meeting, 2023.

[129] ATLAS Collaboration, SALT documentation, Accessed: 2024-08-10 (2020).
URL https://ftag-salt.docs.cern.ch

[130] J. Birk, et al., umami-hep/puma: 0.3.0 (oct 2023). doi:10.5281/zenodo.8401864.
URL https://doi.org/10.5281/zenodo.8401864

[131] J. Barr, et al., Umami: A Python toolkit for jet flavour tagging, Journal of Open Source Software 9 (102) (2024) 5833. doi:10.21105/joss.05833.
URL https://doi.org/10.21105/joss.05833

[132] L. Garren, et al., Monte Carlo Particle Numbering Scheme, https://pdg.lbl.gov/2007/reviews/montecarlorpp.pdf, Accessed: 2023-07-10.

[133] S. Navas, et al., Review of Particle Physics, Phys. Rev. D 110 (3) (2024) 030001. doi:10.1103/PhysRevD.110.030001.