

FACULTÉ DES SCIENCES

DÉPARTEMENT DE PHYSIQUE NUCLÉAIRE ET
CORPUSCULAIRE

THÈSE DE MASTER

Improving diboson tagging efficiency for searches for new physics at the LHC

Auteur

Charlotte WALTREGNY

Superviseurs

Prof. Giuseppe IACOBUCCI

Dr. Steven SCHRAMM

Dr. Roland JANSKY



**UNIVERSITÉ
DE GENÈVE**

2017-2018

Contents

1	INTRODUCTION	4
2	THEORY	5
2.1	The Standard Model	5
2.1.1	Fermions	5
2.1.2	Bosons	6
2.2	Limitations	7
2.3	Physics beyond the Standard Model	9
3	THE ATLAS EXPERIMENT	10
3.1	The LHC	10
3.1.1	Description and characteristics	10
3.1.2	LHC past and future performances	11
3.2	The ATLAS Detector	11
3.2.1	Description and coordinate system	11
3.2.2	Inner detector	13
3.2.3	Calorimeters	14
3.2.4	Muon spectrometer	17
3.2.5	Data systems and analysis	17
4	JETS	19
4.1	Introduction	19
4.2	Jet reconstruction	21
4.2.1	Inputs	21
4.2.2	Anti- k_T algorithm	25
4.3	Jet grooming	28
4.4	Large- R jet calibration	28
4.5	Jet substructure variables	29
4.5.1	Mass	29
4.5.2	D_2	31
4.5.3	Angularity	33
4.5.4	KtDR	35
4.5.5	Conclusions	37

5	DIBOSON TAGGING	38
5.1	Motivation	38
5.2	Two-variable tagger	40
5.2.1	Selections on mass and D_2	40
5.2.2	Comparison using different inputs	42
5.3	Three-variable tagger	44
5.3.1	Selections on Angularity and KtDR	44
5.3.2	Comparison using different inputs	46
5.4	Summary	47
5.4.1	Significance gain using a third variable	47
5.4.2	Significance gain using TCCs and a third variable	48
6	CONCLUSION	49

CHAPTER 1

INTRODUCTION

The LHC is probing new physics at an unprecedented energy of $\sqrt{s} = 13$ TeV, opening a world of potential discoveries. One of them is the existence of a hypothetical particle which would decay in two electroweak gauge bosons. To be able to identify the bosons, we look at the final products of the process, that is the pair of quarks they each produce. However, quarks cannot be observed as single objects so they create cones of particles called jets and these objects are the ones that leave a signal in the ATLAS detector. Therefore, a low-energy boson is reconstructed as two small radius hadronic jets, one for each quark. However, bosons produced by a high energy particle become highly boosted and consequently their decay products are extremely collimated. They therefore cannot be reconstructed as two separate jets anymore but rather as a single large radius jet englobing the two quarks. This is where the jet substructure techniques become essential, to be able to get information on the parent particle of interest. These techniques are used to resolve the constituents present within the jet and therefore give additional information on the nature of the particle emitting it. The aim of this thesis is to improve these jet substructure techniques by combining low-correlated jet substructure variables in the goal of improving the identification of boosted vector bosons. A focus is also given on the usage of TrackCaloClusters instead of the traditionally used topo-clusters for the improvement of the tagging efficiency.

The thesis is organised as follows. First, the Standard Model and its limitations are presented in Chapter 2, as well as theories Beyond the Standard Model that could fill the shortcomings encountered by particle physics. In Chapter 3, the ATLAS detector and the LHC are described. Jets and jet substructure techniques are explained in Chapter 4, along with the reconstruction algorithm used, grooming techniques and calibration applied. In Chapter 5, the performance of the jet substructure technique making use of the traditional two-variable tagger are compared with the performance obtained from the three-variable tagger. In this chapter, a comparison of the performance when using TrackCaloClusters instead of topo-clusters as inputs to the jet algorithm is also presented. The conclusion of the thesis is drawn in Chapter 6.

2.1 The Standard Model

The Standard Model (SM) is the theory describing our current understanding of particle physics. It provides a successful description of the fundamental constituents of the Universe, the elementary particles, and of the electromagnetic, weak and strong interactions between them [1]. The SM was developed over many years in the second half of the 20th century and is considered to be a triumph of modern physics as it has explained almost all experimental results and as its predictions have been confirmed for a wide variety of phenomena. Elementary particles in the SM are classified between two main categories: the *fermions*, constituents of matter and the *bosons*, mediators of forces. Figure 2.1 provides a summary of the classification of elementary particles in the SM and consequently of the discussion which will follow in the rest of the section.

2.1.1 Fermions

Fermions have, by definition, half-integer spin value and can be divided into two families called *quarks* and *leptons*, as represented in Figure 2.1. Together and in pairs, quarks and leptons are grouped in three generations, ranging from the first that makes everyday matter to the second and third describing heavier and less stable counterparts. The electron and the electron neutrino are leptons and, together with the up quark and the down quark, they form the *first generation*. The *second generation* is formed by the two leptons called muon and muon neutrino and by the strange quark and the charm quark. The *third generation* is formed by the remaining two leptons, the tau and the tau neutrino, and by the top and bottom quarks. The quarks have a *flavour*, a colour charge, a mass and an electric charge. The leptons have a *leptonic number* but only the electron, muon and tau have a mass and an electric charge. The neutrinos have no electric charge and, in the SM, they are considered to be massless (even though they do have a very small mass [9]).

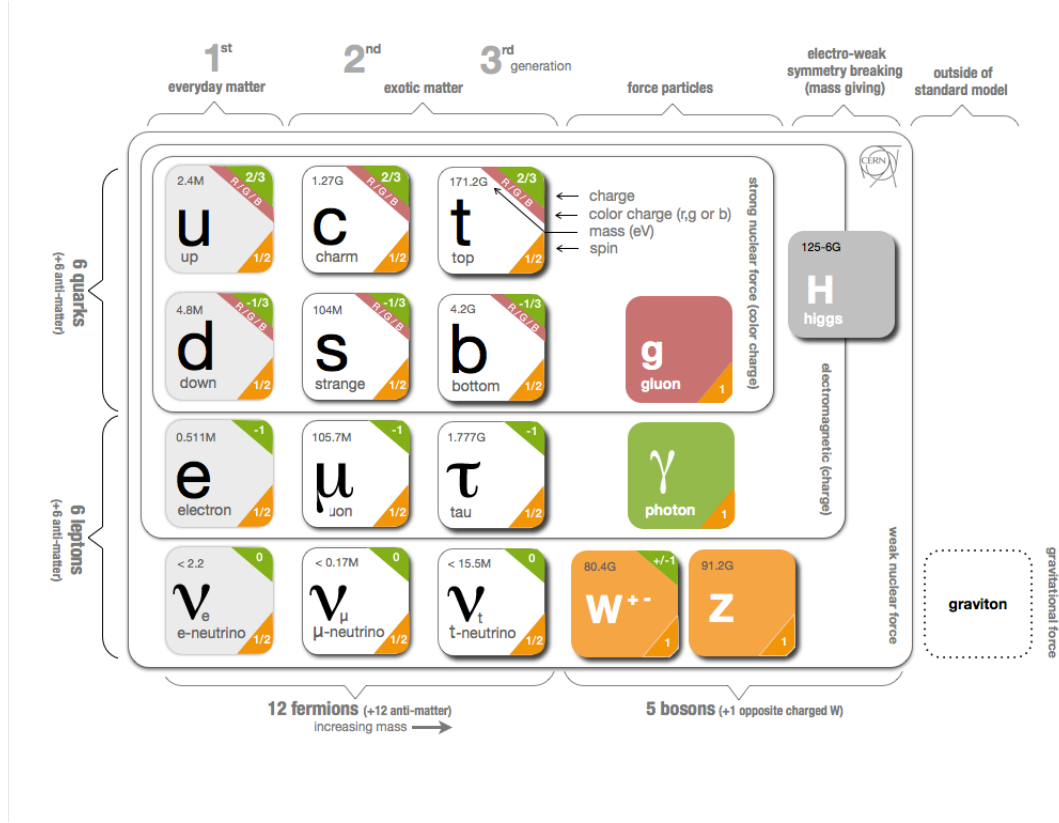


Figure 2.1: Classification of particles in the Standard Model [2].

2.1.2 Bosons

The elementary particles interact with each other through the means of four fundamental forces: *gravity*, *electromagnetism*, *strong force* and *weak force*. The gravitational force can be neglected when talking about particle interactions as it is extremely small compared to the other three at the LHC energy scale. The bosons have spin integer value and are the carriers of the remaining three fundamental forces. The photon, the gluon, the W^\pm and the Z are the four *gauge bosons* and are spin-1. The *Higgs boson* is spin-0 and is the last element of the SM. Photons and gluons carry respectively the electromagnetic force and the strong interaction, and are both massless. The weak-charged current interaction is mediated by two massive bosons, W^+ and W^- , and is responsible for the nuclear fusion occurring in the core of stars [1]. The weak-neutral current interaction is mediated by the Z boson and together with W^\pm , are the carriers of the weak force. Figure 2.2 summarizes all possible interactions between bosons and fermions that are allowed by the SM.

Interactions between particles are extremely important as the properties of the fermions are defined by the set of interactions they feel. The quarks interact via the strong interaction, the weak interaction and the electromagnetic force whereas the leptons only interact via the weak and electromagnetic forces. Consequently, apart from neutrinos, all fermions experience the electromagnetic interaction described in a fundamental theory called Quantum Electrodynamics (QED). Among these fermions, only the quarks feel the strong force described in a similar fundamental theory called Quantum Chromodynamics (QCD). Even though QED and QCD are

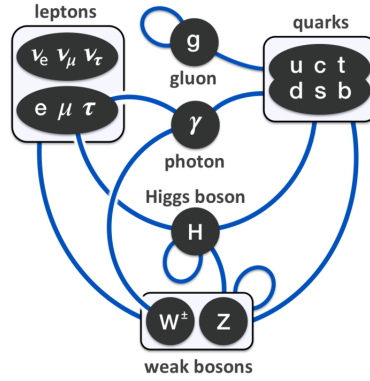


Figure 2.2: Interactions between elementary particles in the Standard Model [2].

similar as QED served as a model for the construction of QCD, there is one crucial difference. The photon (the force-carrying particle in QED) does not possess an electric charge and, consequently, photons cannot interact with other photons. On the contrary, the gluon (equivalent force-carrying particle in QCD) possess a color charge (the QCD equivalent of the electric charge) and gluons therefore interact with each other. This has a fundamental effect: quarks cannot exist by themselves and they consequently group themselves into structures called *hadrons*, such as protons or neutrons. This process is known as *color confinement* and will be further discussed in section 4.1. Out of an attempt to build a gauge theory for the weak interaction similar to QED, Glashow, Salam and Weinberg successfully unified weak and electromagnetic forces under the electroweak theory [3, 4, 5]. Consequently, the Standard Model is the quantum field theory grouping electroweak theory and QCD which together describe the behavior of all known elementary particles. The Higgs boson, which is the only scalar particle (spin-0), completes this beautiful picture by providing the mechanism through which all fundamental particles obtain their masses [6, 7].

2.2 Limitations

The Standard Model is revolutionary and has demonstrated numerous times its powerful predictions such as the existence of the gluon, the W boson, the Z boson, the top, bottom and charm quarks and inevitably the Higgs boson. The SM also predicted precisely properties of some of these particles (mass, spin, charge...) before they were even observed. The question is: if we have such a powerful theory answering most of our questions, why do we continue to study particle physics?

It is the "most of" that bothers us here. The SM is only able to explain three out of the four fundamental forces present in our Universe. There is no equivalent particle to the photon of the electromagnetic force or the gluon of the strong force for the gravitational force, the so-called *graviton*, in the SM.

Furthermore, it fails to explain why gravity is so weak compared to the other four fundamental forces, the weak force being 10^{32} times stronger than the gravitational one. Going into more details, this question leads to another one: why is the Higgs mass so small? In the SM, the values of the particle masses are computed taking

into account quantum corrections. When doing so for the Higgs boson, these corrections become very large because the Higgs mass m_H receives contributions from every particles coupling to the Higgs field. If m_H is the observable Higgs mass, m_H^0 the bare Higgs mass and Δm_H the quantum corrections to the mass, the relation linking them is given by

$$m_H^2 = (m_H^0)^2 + \Delta m_H^2$$

with $\Delta m_H^2 \approx 10^{32}$ if the SM is effectively the theory describing physics at the Planck scale. As the Higgs mass is 125 GeV, we have $m_H^2 = 15625 \approx 10^5$. This means that, in order to obtain the observable value of the mass of the Higgs, the raw value value has to be extremely *fine-tuned*. In other words, $(m_H^0)^2$ and Δm_H^2 should be the exact same 32 digits-long number with only the last 5 digits changing to be able to obtain the Higgs mass, which seems very strange. Figure 2.3 illustrates this example with $(m_H^0)^2$ in red, Δm_H^2 in blue and m_H^2 in green.

$$\begin{array}{r} 12345678901234567890123456789012 \\ - \\ 12345678901234567890123456773387 \\ = \\ 15625 \\ = \\ (125)^2 \end{array}$$

Figure 2.3: Example illustrating the fine-tuning of the raw value of the Higgs mass if SM was the theory describing nature at the Planck scale [8].

This problem is known as the *hierarchy problem* where the answer provided by the SM seems completely unnatural.

Another limitation already mentioned briefly at the beginning of the chapter is the massless description of neutrinos. Indeed, the observation of neutrino oscillations by the Super-Kamiokande Collaboration [9] showed that neutrinos do possess a mass that is not predicted by the SM. This mass induces a CP violating phase in the neutrino mixing matrix which could lead to explanations for the matter-antimatter asymmetry.

Antimatter is the exact copy of matter but with opposite charge. For example, hydrogen is formed by a proton and an electron, and anti-hydrogen by an anti-proton (negatively charged proton) and an anti-electron (positively charged electron) [10]. During the Big-Bang, the same amount of matter and antimatter was created but now our visible Universe is composed almost entirely of matter. Understanding where this asymmetry between matter and antimatter comes from is one of the main concerns of particle physics.

Astrophysical observations have shown that visible matter accounts for only 5% of the total matter present in our Universe [11], that is stars, galaxies, planets etc. The remaining 95% is what is referred to as *dark matter* and *dark energy*. They are called that way because they don't interact with light and are thus invisible to the electromagnetic spectrum and therefore to us. Only their effects can be inferred

from indirect observations such as the acceleration of the expansion of the Universe. In the SM, no particles have the properties to correspond to dark matter nor dark energy.

2.3 Physics beyond the Standard Model

Due to the limitations faced by the SM, new theories have been developed and are grouped under the name "theories of physics *Beyond the Standard Model* (BSM)". These theories aim at explaining the issues encountered in the SM and some try to answer fundamental questions such as "can the forces be unified?" Indeed, it has been noted that the strength of the three forces in the SM (defined by a number called *coupling constant* and varying with energy) are similar and that these coupling constants may meet at some point in the high-energy scale and thus tend to become one single value. This Grand Unified Theory (GUT) [12, 13] would unify the fundamental interactions of the SM and would be an intermediate step towards the unification of gravity with the other fundamental forces. This GUT, along with other BSM theories, predicts the existence of heavy particles that would decay into vector boson pairs (WW, ZZ, WZ).

The final state of this *diboson* production is the decay of the vector bosons and this thesis concentrates on the fully-hadronic channel, when both vector bosons decay into a pair of quarks. This decay mode has the advantage of having a large branching ratio (BR) of $\text{BR}(W \rightarrow q\bar{q}') \approx 68\%$ and $\text{BR}(Z \rightarrow q\bar{q}) \approx 69\%$ compared to the BR of leptonic decays $\text{BR}(W \rightarrow \ell\nu) \approx 10.8\%$, $\text{BR}(Z \rightarrow \ell\ell) \approx 3.4\%$ (per lepton flavour) and $\text{BR}(Z \rightarrow \nu\nu) \approx 20\%$ (all types combined). However, a large QCD background is produced due to the hadronisation of the quarks forming collimated cones of particles called *jets*. The search for diboson production therefore relies on the suppression of the dijet background using jet substructure techniques. Previous studies in the fully-hadronic channel were performed at $\sqrt{s} = 8$ TeV [16] and at $\sqrt{s} = 13$ TeV [17] but these results exclude at the 95% confidence level the presence of new resonances for masses up to 2 TeV for different benchmark scenarios.

This thesis presents the use of multiple jet substructure variables combined with the benefits of TrackCaloClusters in order to further discriminate the background from the signal and consequently improve the tagging procedure.

THE ATLAS EXPERIMENT

3.1 The LHC

3.1.1 Description and characteristics

The Large Hadron Collider (LHC) [20, 21, 22] is a proton-proton collider located at CERN, the European Organization for Nuclear Research. It was built to achieve energies up to $\sqrt{s}=14$ TeV which makes it the world's largest and most powerful particle accelerator ever created. It lies inside a 27km-long tunnel, 100 meters underground below the France-Switzerland border, near Geneva. The aim of the LHC is to test the predictions of the Standard Model, including the search for the Higgs boson, and to go past its limitations by investigating theories Beyond the Standard Model.

The LHC has two different beam pipes where the protons are circulating and four interaction points where the beams meet each other in order to produce the collisions. These are where the main LHC experiments are located: ALICE, ATLAS, CMS and LHCb.

Before being injected inside the LHC, the protons go through four accelerating steps depicted in Figure 3.1:

- LINAC2: After being extracted from hydrogen molecules, the protons are first accelerated by a linear accelerator which takes their energy up to 50 MeV.
- BOOSTER: This is the first ring of the acceleration chain which brings the protons to an energy of 1.4 GeV.
- PROTON SYNCHROTRON: Accelerates the particles to an energy of 26 GeV.
- SUPER PROTON SYNCHROTRON: This is the last step before the injection into the LHC, which brings the energy of the protons to 450 GeV.

The LHC accelerates the particles to the collision beam energy of 4 TeV in 2012 and 6.5 TeV in 2015. Then, thousands of magnets are used to direct and guide the particles inside the pipes. Dipole and quadrupole superconducting electromagnets are used in order to respectively bend and focus the beams.

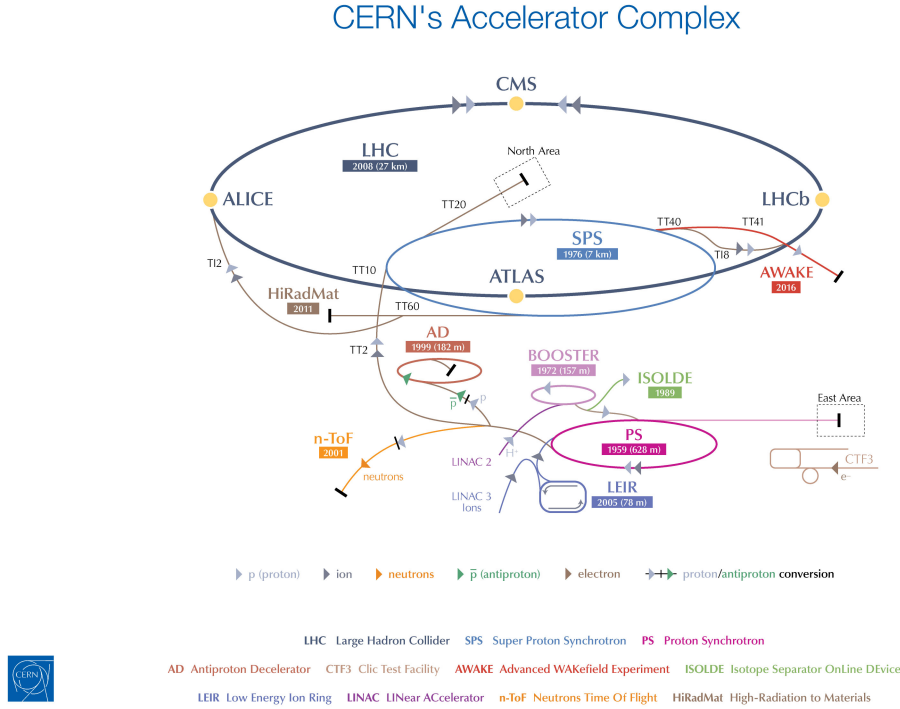


Figure 3.1: CERN's Accelerator Complex [23].

3.1.2 LHC past and future performances

The first beam circulated inside the LHC on 10 September 2008 but nine days later an electrical fault led to the damage of 53 superconducting magnets and inevitably to a year of repairs. In November 2009, the first collision occurred at an energy of 450 GeV soon to be followed, in December 2009, by the record-breaking collision at $\sqrt{s}=2.36$ TeV. Improvements were made in the following years, first reaching $\sqrt{s}=7$ TeV in March 2010 then achieving the maximum performances for Run 1 in 2012 with a collision energy of 8 TeV and instantaneous luminosity of $7.73 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. After a 2 year shutdown for its upgrade, the LHC is now reaching a collision energy of $\sqrt{s}=13$ TeV and an instantaneous luminosity of $5.02 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. At the end of Run 2 in 2018, the LHC will shutdown for a few more years in order to further increase its luminosity for Run 3 [24]. After Run 3, another shutdown is planned beginning of 2024 for the HL-LHC upgrade which plans to increase the total integrated luminosity by a factor of 10 [25].

3.2 The ATLAS Detector

3.2.1 Description and coordinate system

The ATLAS (A Toroidal LHC ApparatuS) experiment [26] is one of the two general-purpose experiments at the LHC, the other one being CMS. Both experiments are designed to exploit the full discovery potential of the LHC and to take advantage of the large amount of information the collider delivers. ATLAS is the largest collider

detector ever built by mankind; it is 44 meters long, 25 meters high and weighs about 7000 tons, as heavy as the Eiffel Tower. The experiment is supported by a huge collaboration of more than 3000 scientists from 38 different countries. The layout of the ATLAS detector is depicted in Figure 3.2.

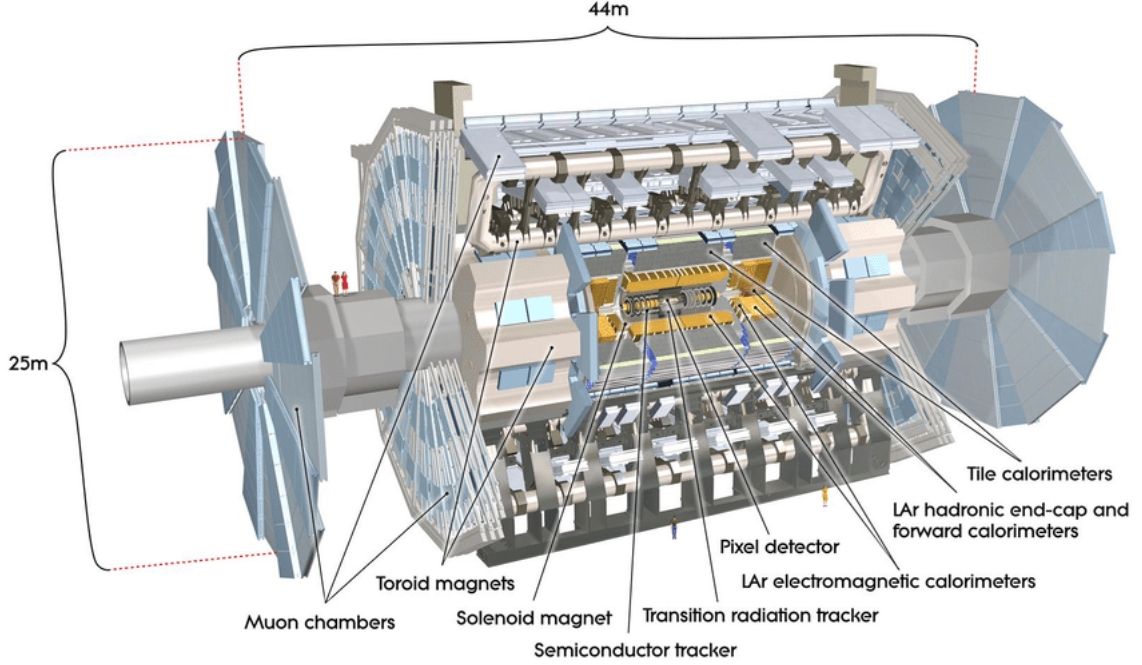


Figure 3.2: Schematic view of the ATLAS detector and its components [26].

The ATLAS detector can be divided into four main parts: the *Inner Detector*, the *calorimeters*, the *Muon Spectrometer* and the *magnet system*. The following sections describe these parts in more details, with a focus on the Inner Detector and the calorimeters due to their importance in jet reconstruction, and therefore in the subject of this thesis.

Before entering the details of the composition of the ATLAS detector, we must first describe the set of coordinates used for the rest of the discussion. The interaction point is set as the origin of the system with the beam direction defining the z axis, the x axis pointing towards the center of the LHC and the y axis pointing upwards. The (R, ϕ, z) coordinate system is chosen instead of the cartesian (x, y, z) one due to the cylindrical geometry of the ATLAS detector. R is defined as the distance from the z axis in the transverse plane $R = \sqrt{x^2 + y^2}$ and ϕ is the azimuthal angle which is the angle from the x axis in the transverse plane. A set of variables also needs to be defined:

- The pseudorapidity η : it is defined as $\eta = -\ln(\tan \frac{\theta}{2})$ where θ is the polar angle measured between the momentum of the particle and the z axis. Rapidity is preferred instead of θ because differences in rapidity are Lorentz invariant under boosts along the longitudinal axis. Pseudorapidity is a good approximation of it and is easier to use in experimental conditions.

- *The transverse momentum p_T* : it is the momentum of the particles in the xy plane and it is defined as $p_T = \sqrt{p_x^2 + p_y^2}$ where p_x and p_y are the momenta of the particles in the x and y directions respectively. The transverse momentum is used because momentum is conserved in the xy plane, that is, the sum of the p_T of all particles is equal to zero before and after the collision.
- *The distance ΔR* : The distance between particles in the η - ϕ plane is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$

3.2.2 Inner detector

The Inner Detector (ID) is the closest detector to the interaction point. It is designed to track charged particles in order to reconstruct the momentum, the vertex of the event and flavour tagging. It has a η coverage of 2.5 and a complete ϕ coverage to limit the risk of particles going out of the interaction region undetected. To extract the momentum of the particle, a 2T superconducting solenoid surrounds the ID. The magnetic field bends the charged particles and their momenta can be obtained from the curvature of their paths. Furthermore, the ID has a high granularity in order to provide both a good momentum and excellent angular resolutions. The proximity of the interaction point also needs to be taken into account as the materials must resist the extreme radiation they are exposed to during the collisions. To accomplish its goals, the ID is made of three sub-detectors: the *Pixel Detector*, the *Semi-Conductor Tracker* (SCT) and the *Transition Radiation Tracker* (TRT) as depicted in Figure 3.3.

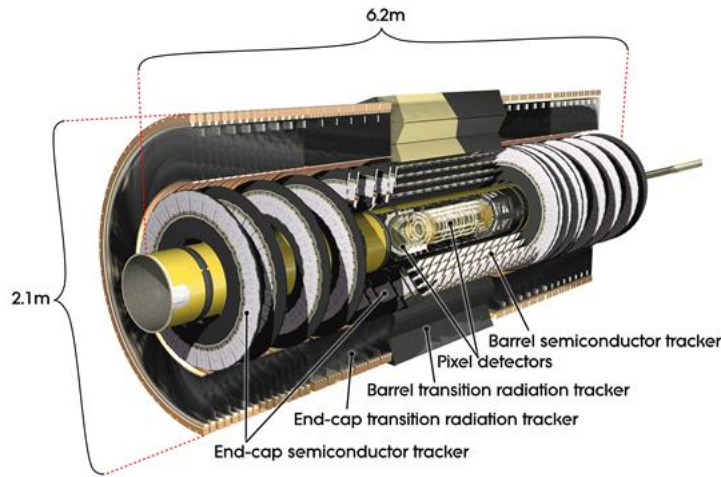


Figure 3.3: Layout of the Inner Detector during Run 1 [26].

The Pixel Detector is a silicon detector composed of three barrel layers and three disks per end-cap close to the interaction point, allowing for a coverage up to $|\eta| < 2.5$. This closeness from the interaction point results in a high exposure to the radiation and therefore in the need for the Pixel Detector to be cooled down to temperatures below -5°C . Its high granularity allows precise measurements necessary for exceptionally accurate tracking and reliable vertex reconstruction.

The SCT is a silicon strip detector outside the Pixel Detector and is composed of four barrel layers and nine disks per end-cap, allowing for a coverage up to $|\eta| < 2.5$ as well.

The TRT is a gaseous detector and is the outermost of the three tracking sub-systems of the ATLAS Inner Detector. It is composed of about 300.000 straws filled with gas mixture and a gold-plated tungsten wire inside. When a particle crosses a straw, it ionises the gas inside. The liberated charges are then collected by applying a high electric field between the wire and the boundary of the straw. In addition to its tracking capabilities, the TRT can also be used for particle identification. Indeed, the gas mixture is sensitive to the number of photons produced which is related to the characteristic of the particle. It allows for a η coverage up to $|\eta| < 2$.

The *Insertable B-Layer* (IBL) was added after Run-1, during the first long shut-down of the LHC, to allow for an improved resolution of the impact parameter and b-tagging performance. It is the fourth and innermost sub-detector of the ID.

3.2.3 Calorimeters

Definition and overall characteristics

Calorimeters are an important part of the detector as they transform the energy deposition of a particle into a measurable quantity. Calorimeters come in two types: *Electromagnetic Calorimeters* are used to measure electrons and photons via their electromagnetic interactions with the materials (bremsstrahlung, pair production etc.) while *Hadronic Calorimeters* are used to measure hadrons via both strong and electromagnetic interactions. ATLAS uses both electromagnetic and hadronic *sampling calorimeters*, meaning that they are made of a succession of layers of active and absorbing materials. When a particle hits the absorbing region, an interaction occurs producing a high quantity of secondary particles. These secondary particles continue to go deep inside the calorimeter hitting succeeding numbers of absorbing plates and producing a cascade of particles called *shower*, until the process eventually stops when the resulting particles do not have sufficient amount of energy to continue the process [27]. As the shower develops, the particles pass through the active or detecting regions where their energy is measured by estimating the amount of ionisation produced when they do. There are two types of showers: electromagnetic showers induced by electrons or photons and hadronic showers induced by hadrons. These showers propagate differently and therefore different calorimeters are needed to identify them, hence the two present in ATLAS. Hadronic showers are much more complex than electromagnetic ones because whereas most of the energy carried by incoming electrons or photons goes into ionisation, a sizeable part of the hadronic shower is carried away by neutrinos which do not interact with the calorimeters. Therefore, for the same amount of energy deposited by an electromagnetic particle or a hadronic one in the calorimeters, a smaller signal will be induced for hadrons. Due to this characteristic, the ATLAS calorimeters are defined as *non-compensating*. This effect has to be taken into account during jet reconstruction, explained in section 4.2.1.

The energy resolution of the calorimeters can be parametrized using

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus b$$

where a is the stochastic term, due to the fluctuations of the physical development of the shower and where b parametrizes the non-uniformity in its response due to the electronic noise of the readout chain [49]. Energy resolution therefore increases with increasing energies.

The rest of this section describes the electromagnetic and hadronic calorimeters of ATLAS in more details.

Electromagnetic Calorimeter

The ATLAS Electromagnetic Calorimeter (ECAL) is a sampling calorimeter which uses lead as its absorbing material and liquid argon (LAr) as its active material due to its linear behaviour and resistance against radiations. The ECAL is divided into a barrel and two end-caps regions (EMEC). The barrel is designed in an accordion shape composed of three layers as shown on Figure 3.4. The first layer has the highest granularity, allowing the electrons and photons to be differentiated from the pions. The second layer is the largest one in the barrel module and is used to collect most of the energy deposited. To finish, the last layer has the lowest granularity and its purpose is to measure the tails of the electromagnetic showers, thus the high-energy ones.

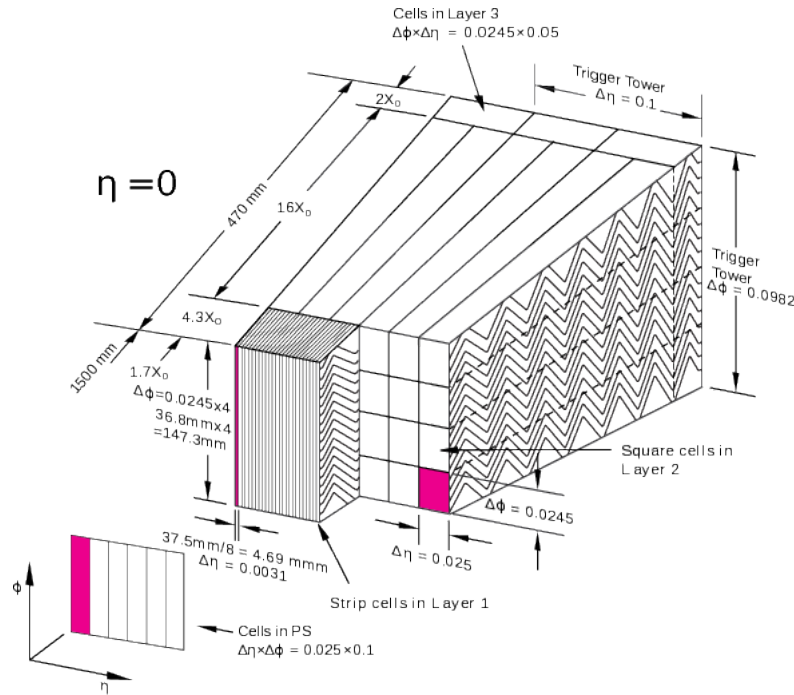


Figure 3.4: Sketch of a barrel module of the EM calorimeter [26].

The accordion shape was chosen so that it provides a full ϕ coverage with no crack regions. The barrel covers an η coverage up to $|\eta| < 1.475$ and has a thickness of at least $22 X_0$, which is the mean path length travelled by an electromagnetic-interacting particle in a material before losing all but $1/e$ of its energy. It is called the *radiation length*. The end-caps cover the regions $1.375 < |\eta| < 3.2$ and their thickness varies as a function of η between 24 and $38 X_0$.

Hadronic Calorimeter

Hadronic showers go deeper inside the calorimeter and so, in order for it to fully contain the shower and that the latter does not reach and damage the Muon Spectrometer, the hadronic calorimeter has to be thicker than the electromagnetic one. The ATLAS Hadronic Calorimeter (HCAL) is a 10 interaction length (λ)-long sampling calorimeter which is divided into three systems: the Hadronic Tile Calorimeter, the Hadronic End-cap Calorimeter (HEC) and the Hadronic Forward Calorimeter as shown in Figure 3.5 depicting the whole calorimeter system.

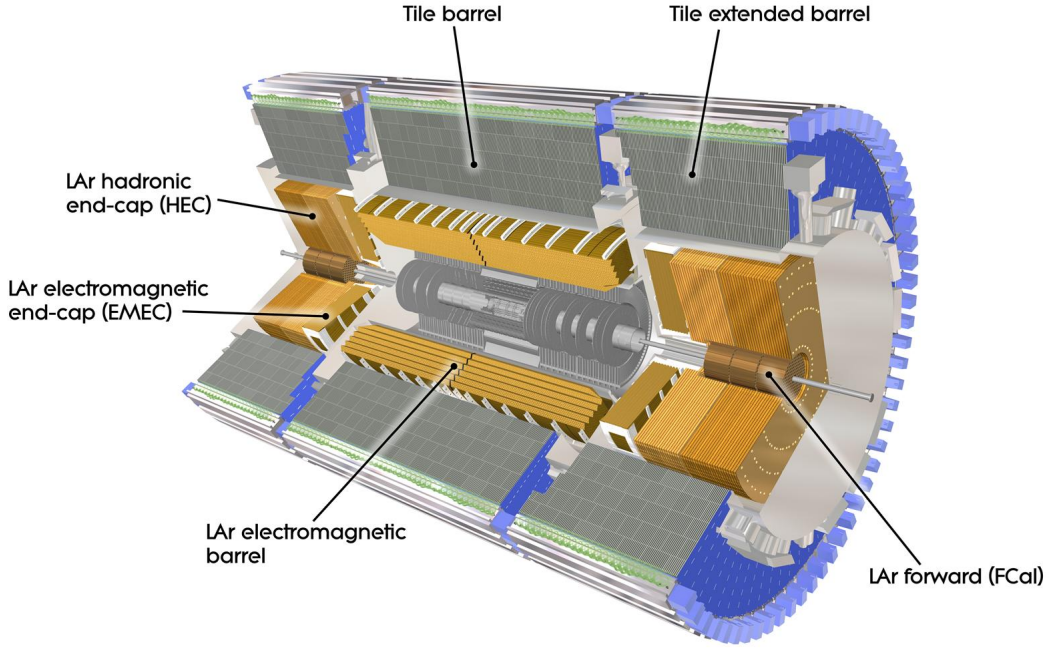


Figure 3.5: ATLAS calorimeter system showing the components of the electromagnetic and hadronic calorimeters [26].

The Hadronic Tile Calorimeter uses steel plates as its absorbing material and plastic scintillators as its active material. It is divided into a central barrel covering up to $|\eta| < 1$ and two extended barrels providing a coverage between $0.8 < |\eta| < 1.7$. The whole Tile Calorimeter is composed of 500.000 scintillator tiles.

The Hadronic End-Cap Calorimeter uses copper plates as its absorbing material and LAr cooled down to -183°C as its active material. It consists of two wheels placed behind the Electromagnetic Calorimeter and provides a η coverage of $1.5 < |\eta| < 3.2$ thus overlapping the extended barrels and the Forward Calorimeter [49].

The Forward Calorimeters

The Forward Calorimeter (FCal) is divided into three layers, all using LAr as their active material but using different absorbing materials. The first layer is used for electromagnetic measurements and the other two for hadronic ones. LAr is used in both cases due to its high radiation tolerance, which is particularly important in the forward region.

3.2.4 Muon spectrometer

The Muon Spectrometer is the outermost subdetector enclosing the Inner Detector and the calorimeters. Its purpose is to track and measure the transverse momentum of muons, because usually these particles traverse the ID and Calorimeters. In order to measure the momentum, superconducting toroid magnets are placed inside the barrel and the two end-cap regions. The location of the barrel layers allows a $|\eta| < 2.7$ and a full ϕ coverage.

3.2.5 Data systems and analysis

Trigger System

The ATLAS detector is designed to observe up to one billion collisions per second, creating a data volume of more than 60TB every second [18]. However, all the collision events are not interesting for physics analysis and in order to reduce this datastream to 1500MB/s, selections have to be applied. The trigger system manages that, dividing the selection process in two steps:

- *The Level-1 trigger*: takes information from the calorimeter and muon detectors in order to make its decision in less than 2 microseconds after the event occurred. It reduces the event rate from 40MHz down to 100kHz.
- *The High-Level trigger*: focuses on specific regions of interest selected by the Level-1 trigger and reduces the event rate down to a few thousand events per second.

Simulation

Monte Carlo (MC) event generators are used to simulate high-energy events produced in particle colliders. They are extremely important as they are used to estimate the SM background, to identify observables necessary for the discrimination between signal and background, to estimate the systematic uncertainties, and more. The simulation of a hard-scattering process can be divided into four steps: the *hard interaction*, the *parton showering*, the *hadronisation* and simulation of *underlying event*. A schematic example of a pp collision is represented in Figure 3.6.

The core of any hadron collision is the hard interaction of the two partons which is generated and computed for the simulation by convoluting the parton distribution functions (PDFs) with the cross section at parton level. As this step involves a large momentum transfer (Q^2), perturbative QCD calculations can be used. The partons are colored objects and so can radiate gluons in the interaction which, in turn, can radiate other particles creating a cascade of emitted particles called parton shower. This process is dominated by low Q^2 and goes on until a certain energy scale is reached where the colored partons recombine to form colorless hadrons in a step called hadronisation (see section 4.1 for further explanations). These hadrons, as well as other heavy objects produced, may not be stable and can decay until a stable configuration is reached. To finish, the interaction between two partons which are not involved in the hard process is simulated. These are the underlying events dominated by low Q^2 and which require the use of phenomenological models as they cannot be calculated by perturbative QCD. Pythia8 [19] is used as the MC

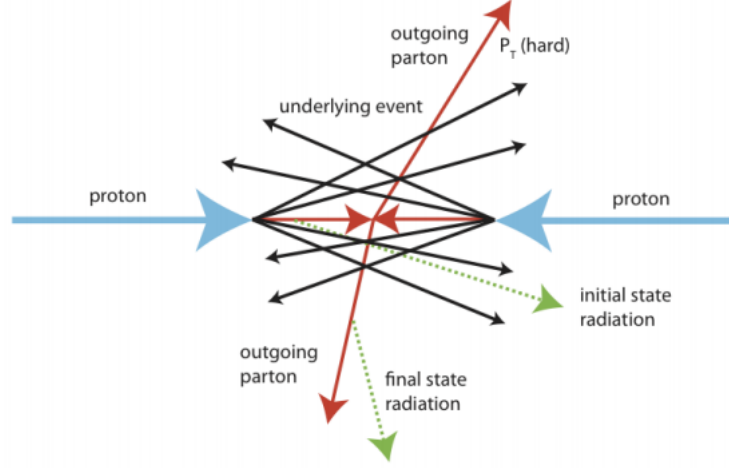


Figure 3.6: Proton-proton collision and underlying events [49].

event generator in this thesis and handles both QCD perturbation calculations at leading order needed for the hard process and the phenomenological models needed for underlying events simulations.

After the event is fully simulated, it is passed through a GEANT4 model of the ATLAS detector to simulate the interaction of the particles with its active and passive materials [50]. Finally, the step of digitisation converts the detector response to the energy deposition in the calorimeters into a signal which will be used during the reconstruction process, where the same reconstruction code is applied to real and simulated data.

4.1 Introduction

As explained in Chapter 2, quarks and gluons cannot exist in isolation and therefore cannot be directly observed as single particles in the detector. Therefore, during a high-energy collision in which they are created, they combine spontaneously with other quarks or antiquarks to form hadrons. This process, which was briefly introduced in section 3.2.6, is called *hadronisation*. Figure 4.1 depicts the process using the example of an interaction producing a quark and antiquark travelling back to back in their centre-of-mass frame. As they separate, the attraction between the two quarks grows stronger and eventually the energy stored in the color field is too high and it becomes more energy-efficient to create new pairs of quarks. This process continues, creating more and more pairs until they have sufficiently low energy to combine themselves into stable hadrons and other colourless objects [1].

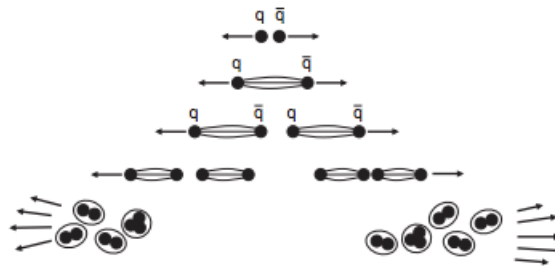


Figure 4.1: Hadronisation process [1].

In this case, two distinguishable group of colourless objects result from the process, one travelling in the initial quark direction and the other in the opposite antiquark direction. These are called *jets of particles*. A jet is therefore a collimated spray of hadrons and other particles. They are what is measured by the detector and therefore used to study the properties of the original quarks.

This thesis focuses on a specific signature, the decay of a heavy particle into two vector bosons (W or Z) that each decay hadronically into a quark-antiquark pair ($X \rightarrow VV \rightarrow q\bar{q}q\bar{q}$). Therefore, what should be observed in the detector is the decay of the two vector bosons into four jets ($X \rightarrow VV \rightarrow jjjj$). However, this is not always the case. Hadronic boson decays can be separated into two regimes [28]:

- *Resolved*: the vector boson is produced with a low transverse momentum, typically $p_T < 160$ GeV. The decay products are therefore resolved into two distinct jets as shown in the top pictures of Figure 4.2.
- *Boosted*: the vector boson is produced with a high transverse momentum p_T , much larger than its mass $m_{W/Z}$. The decay products are therefore highly collimated and begin to overlap, thus they are instead captured into a single *large-radius jet* (defined below) as shown in the bottom pictures of Figure 4.2.

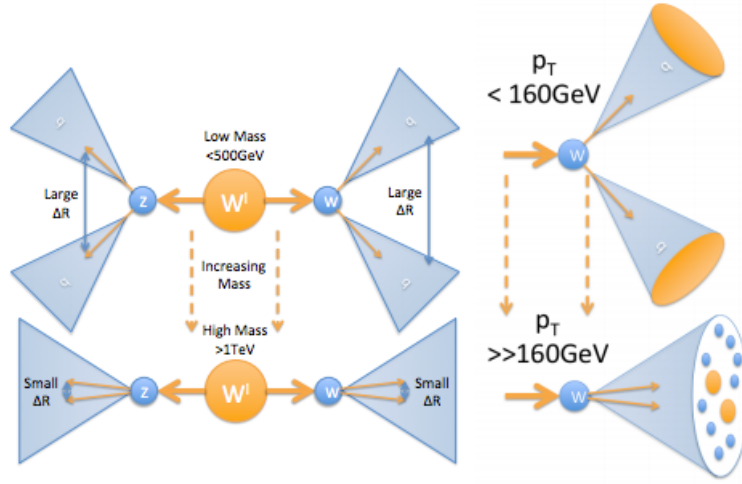


Figure 4.2: Sketch of the decay of a heavy particle into two vector bosons that decay hadronically in the resolved regime (top) and boosted regime (bottom) [28].

Due to the high energy of the collision at the LHC, the decay products of the vector boson are *boosted* and the second regime often applies. The angular separation ΔR_{qq} of the decay products is defined as

$$\Delta R_{qq} = \frac{1}{x(1-x)} \frac{m_{W/Z}}{p_T}$$

with $m_{W/Z}$ the mass of the vector boson, p_T its transverse momentum, x the momentum fraction carried by the first quark and $(1-x)$ the momentum fraction carried by the second quark. In the resolved regime, two typical-sized jets $R=0.4$ can be used to fully resolve the two showers produced by the two quarks. In the boosted regime, however, the two showers start to overlap and it is better to reconstruct the boson as a single large-radius jet $R = 1.0$ (large- R jet), to make sure that all the energy deposited from the hadronic decays is captured within the jet [50]. Due to the high luminosity of the LHC, several collisions can contribute to the detector signals associated with the interesting hard scattering, and therefore low- p_T particles (referred to as soft) can contaminate the event. This phenomenon is called *pile-up*. In addition to soft contributions, high-energy particles originating from other deep

inelastic scatterings (referred to as hard) can result in the production of jets. Jets originating from the boson decays (*bosonic jets* or *W/Z-jets*) therefore need to be distinguished from the ones created by these hard quarks or gluons (QCD-jets or *q/g jets*). The main differences between the two types of jets are, first, that two narrow regions with high energy density corresponding to the two quarks can be identified in bosonic jets whereas only one region corresponding to the hard quark/gluon can be identified in hadronic ones (see Figure 4.3).

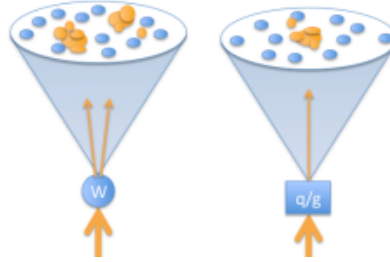


Figure 4.3: Bosonic jet (left) and hadronic jet (right) [28].

Moreover and most importantly, for the *W/Z-jet*, the jet mass originates from the boson and is therefore peaked whereas in the case of QCD-jets, the jet mass comes from the spread of energy deposition and is therefore random [28]. This specific difference is exploited by *jet substructure techniques*, through the use of a new set of observables called *jet substructure variables*, in order to discriminate between signal and background events. These variables are computed from the constituents of the jet and describe the energy distribution inside of it. This concept is of key importance in the subject of this thesis and will be described in depth in section 4.5. But first, to be able to further resolve such substructure properties, *grooming techniques* are applied which aim to identify soft contamination from pile-up and remove them from the jet (see left picture of Figure 4.4). Only after that, the *tagging procedure* using the substructure variables mentioned above can be applied to reject background jets (see right picture of Figure 4.4).

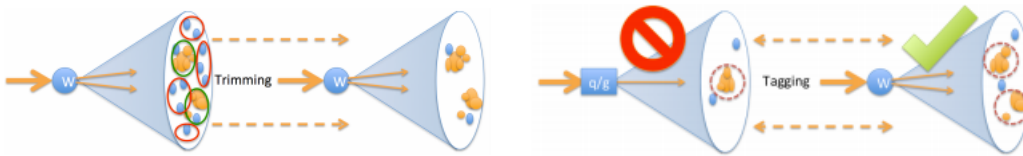


Figure 4.4: Grooming procedure (left) and tagging procedure (right) [28].

4.2 Jet reconstruction

4.2.1 Inputs

Topological clusters

Before discussing grooming and tagging procedures, it is useful to understand how jets resulting by hadrons are reconstructed in ATLAS. That is, how the energy depo-

sition in the calorimeters is then reconstructed as a jet coming from a parent parton. *Topological clusters* (or topo-clusters) [29] are the inputs to the jet reconstruction algorithms in ATLAS. It is a technique combining single calorimeter cells topologically to reconstruct the shower of the hadronisation process. In other words, the idea is to look for a cell which has a significant energy compared to the average noise induced by electronic components and expected pile-up contributions, and to group it with neighbouring cells passing a certain significance threshold. The grouping of the calorimeter cells is therefore based on the cell signal significance ς_{cell}^{EM} defined as the ratio of the cell energy E_{cell}^{EM} and the average cell noise $\sigma_{noise,cell}^{EM}$

$$\varsigma_{cell}^{EM} = \frac{E_{cell}^{EM}}{\sigma_{noise,cell}^{EM}}$$

with quantities measured at the electromagnetic (EM) energy scale, not taking into account the non-compensating character of the calorimeter.

Topo-clusters are formed using a growing-volume algorithm which depends largely on the cell significance ς_{cell}^{EM} and on three parameters S, P, N. It follows a series of steps:

- *Proto-clusters*: if cells have $|\varsigma_{cell}^{EM}| > S$ (by default $S = 4$), they pass the primary seed threshold and therefore have a sufficient significance to be considered as seed cells or *proto-clusters*. The seed cell with the highest significance amongst all the identified seed cells (the "first" proto-cluster) is used by the algorithm as its starting point.
- *Neighbours*: the neighbouring cells satisfying $|\varsigma_{cell}^{EM}| > N$ (by default $N = 2$) are then added to the initial proto-cluster, as well as the iterative neighbours of these cells if they pass this threshold. If one of the cells is, in addition, adjacent to another proto-cluster, both proto-clusters merge to form a single one. Once none can satisfy the $|\varsigma_{cell}^{EM}| > N$ condition anymore, all neighbouring cells passing $|\varsigma_{cell}^{EM}| > P$ (by default $P = 0$) are also added to the proto-cluster.
- *Iterate*: once the first proto-cluster is assembled, the algorithm applies the same procedure to the seed cell with the next-to highest significance, that is the second proto-cluster, then the third, until all seed cells have been analysed.
- *Split*: as already mentioned before, this thesis focuses on boosted vector boson decays, where two quarks can be produced close to one another and can therefore be part of a single topo-cluster. If this is the case, the substructure techniques described in section 4.5 cannot be used anymore. This is why the search for local maxima within topo-clusters is fundamental, as the quarks will leave energy peaks representing the leading and sub-leading quarks. Local maxima are required to happen within the middle layer of the hadronic or electromagnetic calorimeters to provide a reliable energy measurement, to have $E_{cell}^{EM} > 500$ MeV, at least four neighbours and no neighbour with a higher energy [29]. Once a topo-cluster has been identified containing local maxima, it is split between the two.

The 4-vector built from the topo-cluster and that will be used as input in the jet reconstruction algorithm is defined with the mass $m_{cluster}$ set to be zero and the angles $(\eta_{cluster}, \phi_{cluster})$ are derived from the individual cell positions. Finally, the

energy $E_{cluster}$ is the sum of the energy of all the cells contributing to the formation of the topo-cluster.

In section 3.2.2, it was mentioned that the calorimeters were non-compensating and this feature has to be taken into account in the reconstruction process. So far, all quantities were measured at the EM energy scale and therefore, to restore the hadronic energy scale, the topo-cluster is calibrated using a *Local Hadronic Cell Weighting* (LCW or LC) procedure [31]. This method determines the likelihood that the topo-cluster was generated from a hadronic or electromagnetic shower and, depending on the result, applies the relevant calibration. In the case of hadronic topo-clusters, the calibration would account for the non-compensating nature of the calorimeters and in the case of electromagnetic ones, the calibration would adjust the energy losses induced by inactive detector regions.

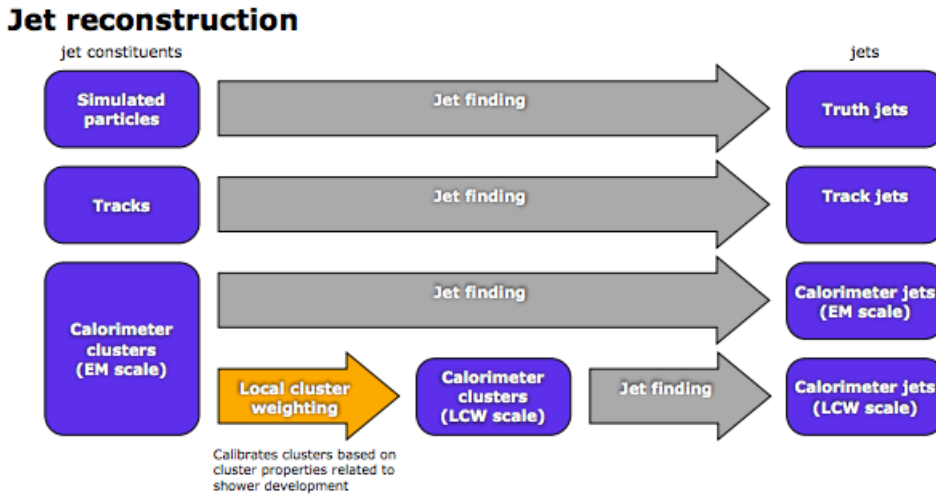


Figure 4.5: ATLAS jet reconstruction [30].

It is worth mentioning the various objects that can be used as inputs for jet reconstruction, as shown in Figure 4.5. In ATLAS, jets are usually built from topo-clusters but any object defined as a 4-vector can be used to reconstruct jets. For example, stable particles with a lifetime of at least 10 ps result in truth jets and charged-particle tracks result in track-jets [32].

Track-CaloClusters

The inputs to jet reconstruction that are of particular interest in this thesis are *TrackCaloClusters*. Track-CaloClusters (or TCCs) [33] are another type of inputs to the jet reconstruction algorithm. The idea behind such objects is to combine the excellent angular resolution of the tracker with the excellent energy resolution of the calorimeter. In the highly boosted regime, where the two quarks resulting from the decay of a vector boson ($V \rightarrow qq$) become highly collimated, the calorimeter lacks the angular resolution to distinguish the two and identifies them as one energy peak, meaning one local maximum. The ability to discriminate between the QCD background and the vector boson signal relies on the ability to resolve the constituents of the jets into separate objects and compute the needed substructure variables. The granularity of the calorimeter therefore becomes a limiting factor in searches for new

physics in highly boosted regimes, even if its energy resolution remains excellent. This is where the tracker can be very beneficial as tracks are less bent with increasing p_T , meaning that the extrapolation of the tracks to the calorimeter becomes more and more accurate and so consequently the angular resolution improves. However, since the p_T of the tracks are computed from the curvature, the energy resolution decreases. This is how the idea of TCCs was born, to use the opposite behaviours of the tracking and calorimeter systems at high energy to create new inputs for jet reconstruction that would have a better angular and energy resolution and would therefore improve jet substructure performances.

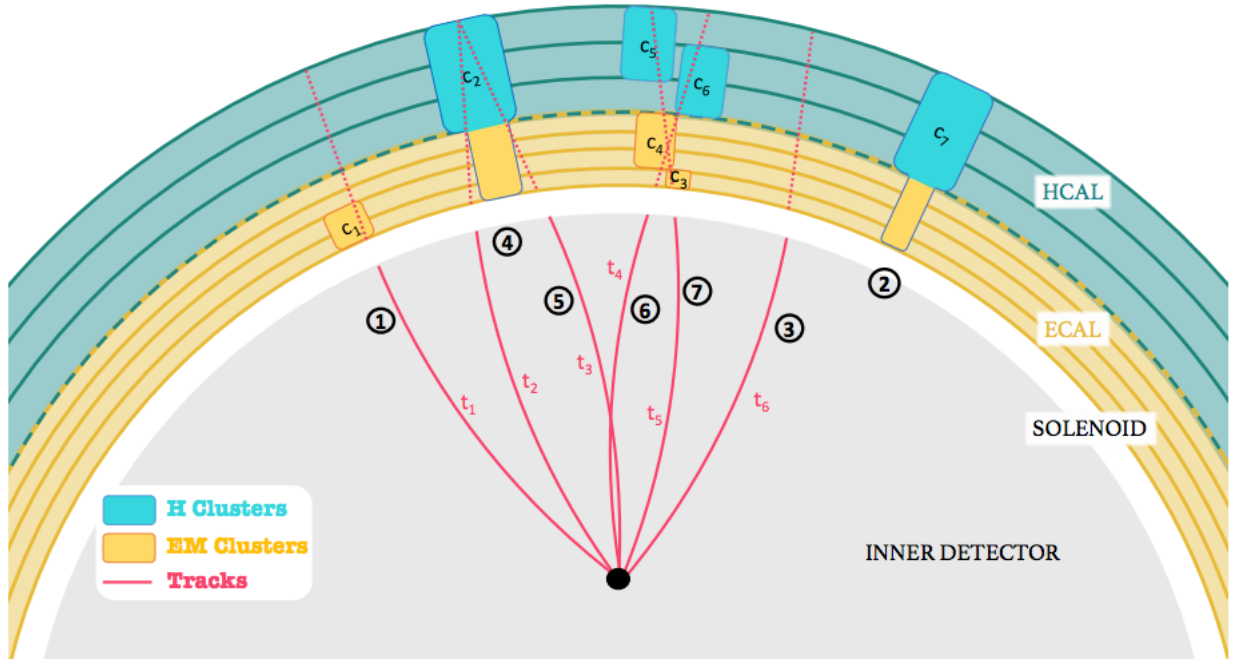


Figure 4.6: Sketch showing seven different TCC objects [33].

TCCs are formed combining tracks and topo-clusters following a specific matching procedure which attempts to match every track to every topo-cluster. First, all tracks with an extrapolation uncertainty σ_{track} larger than the width of the topo-cluster $\sigma_{cluster}$ are eliminated. Otherwise, tracks and clusters are matched if the condition $\Delta R < \sqrt{\sigma_{cluster}^2 + \sigma_{track}^2}$ is satisfied. Once a track is matched with a topo-cluster, the 4-vector is built by taking the energy of the topo-cluster and the direction of the track, with the mass still set to be zero. The concept is quite simple however the implementation is complex because a topo-cluster may not match any tracks and vice versa as shown in ② and ③ of Figure 4.6. Or multiple tracks could be matched to a single cluster but sharing its energy (④, ⑤) or worse, multiple topo-clusters could be matched to multiple tracks (⑥, ⑦). A single algorithm covers all these cases when building the 4-vectors from TCCs. Full details of how TCC objects are built can be found in Reference [33].

4.2.2 Anti- k_T algorithm

A set of rules dictate how particles should be grouped together to form jets, and these are provided by *jet algorithms* [34]. Together with a set of parameters and a recombination scheme defining the momentum to assign to the combined pair of particles, jet algorithms form a *jet definition*. To standardize the definition of jets and therefore make sure that no ambiguity can arise discussing jet results, a group of eminent physicists set a list of properties that should be met by a jet definition [35]:

- Simple to implement in an experimental analysis.
- Simple to implement in the theoretical calculation.
- Defined at any order of perturbation theory.
- Yields finite cross sections at any order of perturbation theory.
- Yields a cross section that is relatively insensitive to hadronisation.

These rules ensure the algorithm to be both *infrared* (IR) and *collinear* safe. An IR-safe algorithm does not change the number of reconstructed jets of the event in the case of soft particle emission, as shown in Figure 4.7

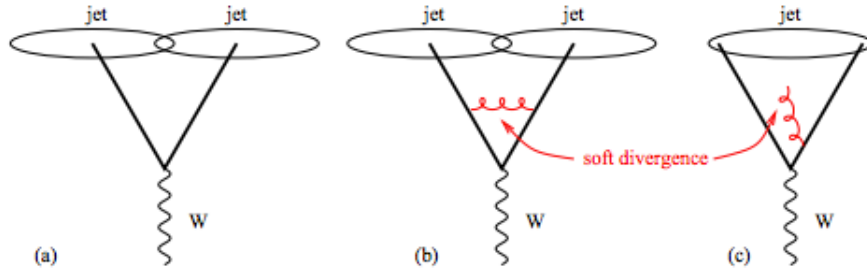


Figure 4.7: Sketch of the hadronic decay of a W boson (a) and the reconstruction of the event using a IR-safe jet algorithm (b) and a IR-unsafe algorithm (c) in the case of the emission of a soft gluon [34].

A collinear-safe algorithm ensures stability in case of collinear splitting, such as the fragmentation of a gluon into two quarks. This concept is of key importance as this process occurs very often in hadronic showers and can lead to a misreconstruction of the event. An example is provided in Figure 4.8 where the height of the vertical lines are proportional to the p_T of the particle and the horizontal line represents the parton rapidity. Figure 4.8 represents the problem when using a collinear-unsafe algorithm: in (d), due to the split of the hardest object into two softer ones, the starting point of the algorithm (the seed) is no longer the hardest object (quark in the middle) and the new hardest parton becomes the quark on the left, leading to a wrong number of reconstructed jets.

There are two types of IR- and collinear-safe algorithms: the *cone-type* algorithms and the *sequential reconstruction* algorithms. The algorithm used in this

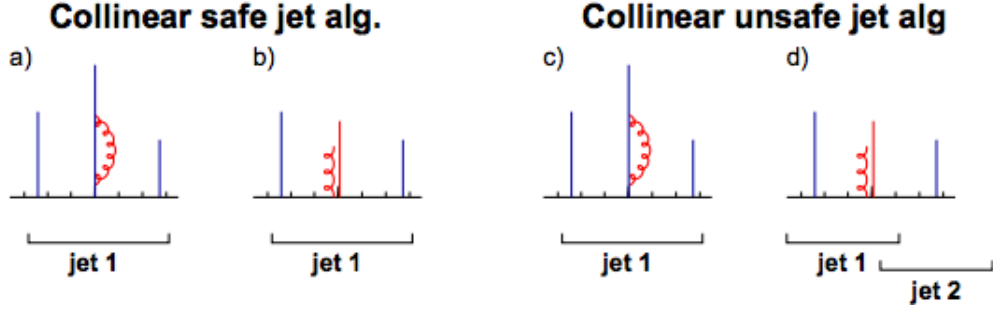


Figure 4.8: Comparison of collinear-safe (left) and collinear-unsafe (right) algorithms [34].

thesis is a sequential recombination algorithm called the *anti- k_T* algorithm [36] and is part of the k_T -family described below. Sequential reconstruction algorithms group the inputs of the reconstruction (topo-clusters, truth or track particles etc.) depending on their distance to each other, combining two particles together at each step [51]. Two different distances are defined, d_{ij} which is the distance between the input i and the input j and d_{iB} which is the distance between the particle i and the beam B . The algorithm identifies the smallest distance there is, if $d_{ij} < d_{iB}$, then particles i and j are combined but, if $d_{ij} > d_{iB}$ for all the entities j , then i is identified as jet and removed from the list of inputs. The algorithm proceeds until no input is left. In the k_T family, the distances are defined as follow

$$d_{ij}^2 = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = p_{Ti}^{2p}$$

where ΔR_{ij}^2 is the angular distance between the two entities i and j , R the radius parameter that enters in the jet definition regulating the size of the jets and p_{Ti} , p_{Tj} respectively the transverse momentum of inputs i and j . One can notice an additional parameter entering the reconstruction algorithm: the parameter p which defines how the characteristics of the entities are combined [50]. The different algorithms present in the k_T -family depend on the value taken by p :

- k_T algorithm: $p = 1$
- Cambridge/Aachen (C/A) algorithm: $p = 0$
- Anti- k_T algorithm: $p = -1$

Figure 4.9 shows the same event reconstructed using the different algorithms. Anti- k_T is the default algorithm used in ATLAS due to the regular circular shapes it provides, making it easier to compute areas and therefore allowing more straightforward jet calibration and pile-up removal [49]. However, the anti- k_T algorithm immediately clusters soft particles to hard ones and so no substructure information can be obtained, as opposed to the k_T algorithm. This is due to the fact that the anti- k_T algorithm follows the development of the shower of the original parton

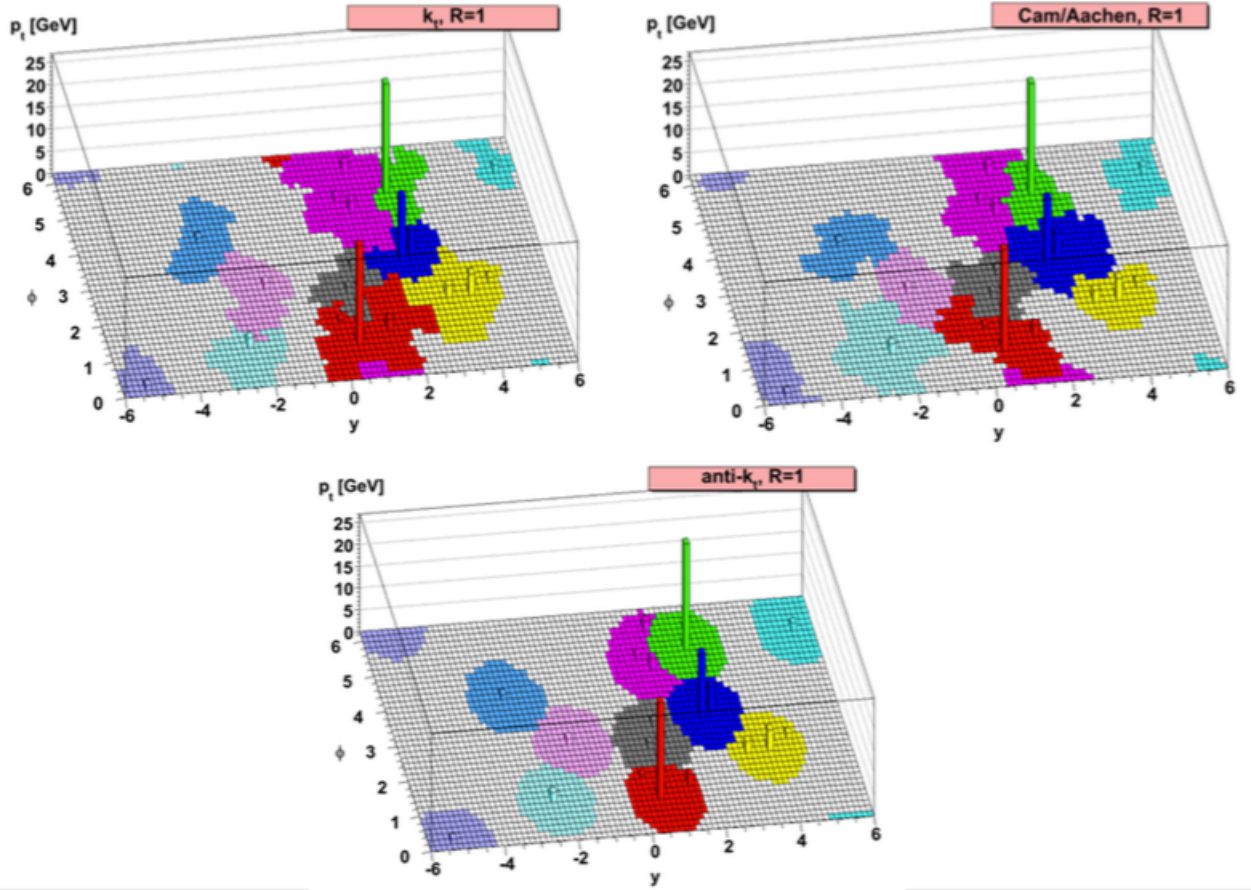


Figure 4.9: Graphs showing how the same event is reconstructed using the k_T algorithm (top left), C/A algorithm (top right) and anti- k_T algorithm (bottom) [36].

whereas the k_T one goes back the showering process, making it possible to remove the last clustering step to extract the hardest constituents of the jets. Once the anti- k_T algorithm has been applied, the irregularities in the shapes of the k_T and C/A algorithms can be used to extract spatial and kinematic substructure information.

4.3 Jet grooming

Particles coming from other collisions, which are not the hard-scatter collision of interest, can contribute to the signal. This is the so-called pile-up mentioned in the introduction of the chapter. During the reconstruction process, these particles might be taken into account and be clustered inside the jets, altering their internal properties. To identify these contributions and to remove them, grooming techniques are applied. There are multiple types of jet grooming are used in ATLAS. The *trimming technique* [37] is the most common, and is used to clean the jets needed for the analysis of this thesis. The trimming algorithm illustrated in Figure 4.10 proceeds as follows [37]:

- Reconstruct the jet of the event (seed jet) using the anti- k_T algorithm with radius parameter $R = 1$.
- Recluster the formed jet into k_T subjets with parameter $R_{sub} = 0.2$.
- Discard the subjet i from the seed jet if it satisfies $p_{Ti}/p_T^{jet} < f_{cut}$ with p_{Ti} and p_T^{jet} respectively the transverse momenta of subjet i and the seed jet. f_{cut} is another parameter of the algorithm, where ATLAS uses 5%.
- Assemble the remaining subjets into the cleaned jet.

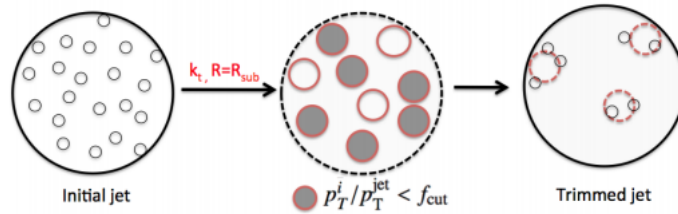


Figure 4.10: Illustration of the jet trimming procedure [32].

4.4 Large- R jet calibration

Jet calibration is of key importance towards the proper analysis of a hard-scattering event because several effects are not taken into account during the reconstruction process. This includes pile-up effects (handled by the trimming procedure described above), the non-compensating character of the calorimeter, differences between the data and MC samples, and similar. [38] Large- R jets are calibrated according to the calibration chain illustrated in Figure 4.11.



Figure 4.11: Sketch of the calibration chain for large- R jets in ATLAS [38].

The calibration of reconstructed large- R jets includes energy and mass calibrations which calibrate the jet energy scale, JES, and the mass scale, JMS, to that of the truth jets (as opposed to the reconstructed jets). In practice, the calibration factors are extracted using the energy response $\mathcal{R}_E = \frac{E_{reco}}{E_{true}}$ which represents the inverse of the calibration function [31]. Since energy and mass are not independent, the JES calibration is first applied. After that, the mass response $\mathcal{R}_m = \frac{m_{reco}}{m_{true}}$ is derived as the inverse of the calibration function. These factors are derived from MC information but, as the latter do not perfectly represent the data, *in situ* corrections can be applied, accounting for these differences and reducing JES and JMS related uncertainties. This completes the current calibration procedure.

4.5 Jet substructure variables

The quarks resulting from the decay of a boosted W or Z boson can be merged into a single jet during the reconstruction process. Jet substructure techniques are used to discriminate between signal and background events within a single jet by characterising the energy distribution inside it. The knowledge of how the energy is distributed within the jet can resolve the two-prong structure of bosonic jets or the one-prong structure of hadronic ones, therefore telling them apart [50]. The substructure variables used in this thesis are described below. Mass and D_2 are used in the standard ATLAS two variables tagger for W identification as this combination of variables was found to give the best background rejection [39]. The goal of this thesis is to create an improved W-tagger by adding a third variable to mass and D_2 resulting in a three variables tagger. This new tagger should outperform the two variables tagger, and Angularity and KtDR are promising variables.

4.5.1 Mass

Introduction

Mass is the most intuitive discriminant because W/Z -jets peak around 80 or 91 GeV (respectively for the W or Z boson) while QCD-jets have a mass distribution, primarily distributed at lower values [51]. Figure 4.12 illustrates these distributions and their differences. The huge separation between the signal and background plots gives an indication of how powerful the mass is as a jet substructure variable.

The jet mass is derived from the sums of the energy and p_T of the constituents i of the jet J , that is

$$m = \sqrt{(\sum_{i \in J} E_i)^2 - (\sum_{i \in J} \vec{p}_i)^2}$$

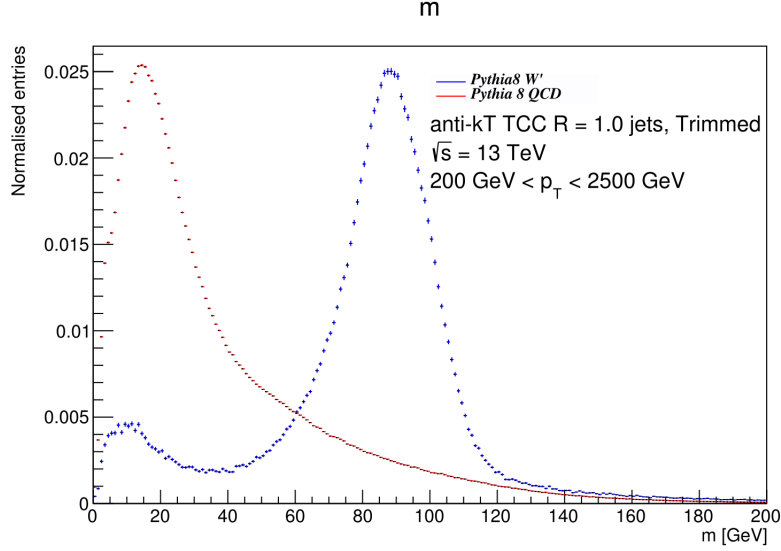


Figure 4.12: Comparison between the mass distributions of QCD and W jets in W' events.

As it depends on the constituents, it also depends on how the jet was built and therefore on the inputs to the jet algorithm. Therefore the mass of calorimeter jets, m_{calo} , is derived from topo-cluster constituents. However, the LHC reached unprecedented energies during Run-2 and the quarks produced in the decay of massive vector bosons become so collimated that their angular distance is comparable with the granularity of the calorimeter. Consequently, a new definition of the mass was introduced, the track assisted jet mass m_{TA} , to make use of the good angular resolution of the tracker at high p_T [40]. The mass of calorimeter large- R jets is now defined as the linear combination between the traditional calorimeter jet mass definition m_{calo} and m_{TA} , and is called the combined mass m_{comb}

$$m_{TA} = m_{tracks} \times \frac{p_T^{calo}}{p_T^{tracks}}$$

$$m_{comb} = m_{calo} \frac{\sigma_{m_{calo}}^{-2}}{\sigma_{m_{calo}}^{-2} + \sigma_{m_{TA}}^{-2}} + m_{TA} \frac{\sigma_{m_{TA}}^{-2}}{\sigma_{m_{calo}}^{-2} + \sigma_{m_{TA}}^{-2}}$$

with $\sigma_{m_{calo}}$ the calorimeter jet mass resolution, $\sigma_{m_{TA}}$ the track assisted jet mass resolution, p_T^{calo} and p_T^{tracks} respectively the transverse momenta of the calorimeter jet and of the four-vector sum of the tracks associated to the large- R jet with m_{tracks} the mass associated to the latter. The combined mass efficiently makes use of the excellent energy resolution provided by the calorimeters at high p_T and by the tracker at low p_T . As mass strongly depends on the energy of the constituents of the jets and of their angular spread, the resolution of both quantities will consequently affect the mass resolution. A comparison of the performances to resolve the mass using topo-clusters and TCCs is presented below.

Mass resolution

It was mentioned in section 4.2.2 that TCCs were created to improve jet substructure performances, especially at high p_T , when the angular resolution of the calorime-

ters becomes a limiting factor in the identification process. The mass resolution of TCC-jets is therefore expected to outperform the one obtained from topo-clusters for $W' \rightarrow WZ \rightarrow qqqq$ samples. Their performances are quantified using the width of the *mass response* ($R = m_{reco}/m_{truth}$), a narrower response indicating an improvement of the resolution. A specific definition of the interquantile range (IQR), which is a measure of statistical dispersion, is used to compute this width. If the distribution was Gaussian, this would correspond to a Gaussian half-width. It is defined as

$$IQR = \frac{1}{2} \frac{Q_{84}(R) - Q_{16}(R)}{Q_{50}(R)}$$

and corresponds to the standard half of a 68% window for the mass resolution used for its measurement in ATLAS [33]. R refers to the the mass response and Q_x corresponds to the $x\%$ quantile boundary. Figure 4.13 shows the mass resolution of TCCs and topo-clusters computed using these definitions. From this plot, we can see that the mass resolution of TCC-jets is superior to the resolution of the combined mass of topo-cluster jets one for truth values of p_T above 2 TeV but is slightly worse below that. This is because the mass of a jet depends more on energy scale than angle until the angle becomes zero.

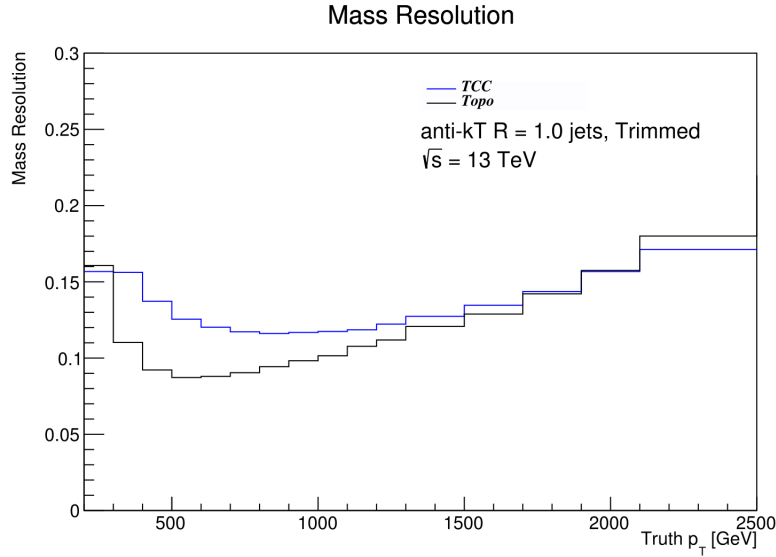


Figure 4.13: Comparison of the mass resolution for jets reconstructed using TCCs and topo-clusters in Pythia8 W' samples.

4.5.2 D_2

Introduction

D_2 , or more accurately D_2^β [41], is a dimensionless ratio of energy correlation functions (ECF) which relate to them by

$$D_2^\beta = \frac{ECF(3,\beta)ECF(1,\beta)^3}{ECF(2,\beta)^3}$$

with the first four N-point $ECF(N, \beta)$ defined as

$$\begin{aligned}
 ECF(0, \beta) &= 1 \\
 ECF(1, \beta) &= \sum_{i \in J} p_{Ti} \\
 ECF(2, \beta) &= \sum_{i < j \in J} p_{Ti} p_{Tj} (R_{ij})^\beta \\
 ECF(3, \beta) &= \sum_{i < j < k \in J} p_{Ti} p_{Tj} p_{Tk} (R_{ij} R_{ik} R_{jk})^\beta
 \end{aligned}$$

These variables run over all particles i in the jet J and depend only on their p_T and on the Euclidean distance between them R_{ij} [42]. β is a free parameter called the angular exponent that can be adjusted depending on the mass of the resonance being studied, in this case the resonance being boosted W/Z bosons. D_2^β is particularly useful for the identification of W/Z -jets because of its ability to identify $N = 2$ dense cores of energy in the jet, hence its name. Small values of D_2^β are characteristics of signal jets whereas large values of D_2^β correspond to background jets [41]. The discrimination power can be further improved by applying a mass window requirement which will also make the variable infrared and collinear safe. Figure 4.14 shows the differences of the $D_2^{\beta=1}$ variable for the two types of jets with and without the mass window restriction. It can be seen that the background jet distribution is more peaked and shifted to the right after the cut on the mass has been applied, which increases the separation between signal and dijet events. $\beta = 1$ is chosen because it better separates between W/Z and QCD jets after a mass cut, so $D_2^{\beta=1}$ will be referred to as D_2 in the rest of the thesis.

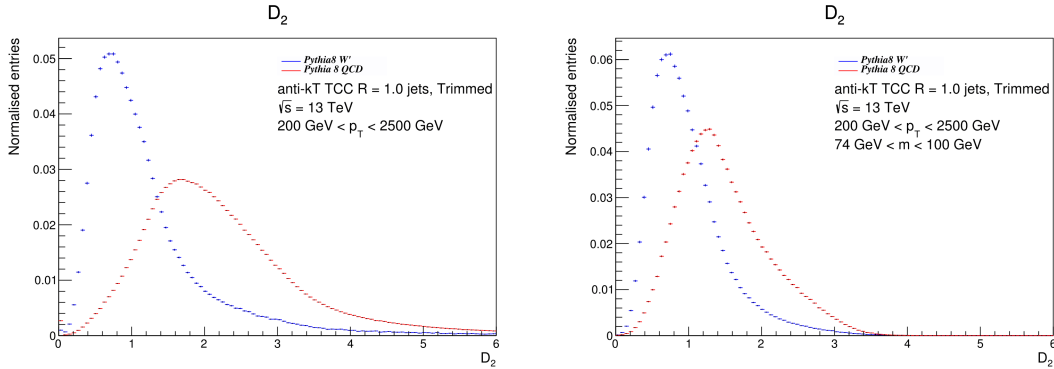


Figure 4.14: Comparison between the D_2 distributions of QCD and W jets, before mass window requirement (left) and after a mass cut of $60 < m < 100$ GeV has been applied (right).

D_2 resolution

A comparison of the D_2 resolution of jets reconstructed using TCCs and topoclusters is performed in this section. As can be seen from its definition, D_2 is

extremely dependent on the angular distance of the inputs to the jet reconstruction. TCC performance is therefore expected to be better than topo-cluster as the TCC definition improves the angular resolution of the inputs to the jet reconstruction [33]. Here, the performance is quantified by comparing the width of the D_2 residuals ($Res = D_2^{reco} - D_2^{truth}$), useful for variables for which the truth value can be 0 and therefore when the response cannot be used. The widths are here computed using the classic definition of IQR:

$$IQR = \frac{1}{2}(Q_{75}(Res) - Q_{25}(Res))$$

which corresponds to the half of a 50% window for the D_2 resolution. Figure 4.15 compares the D_2 resolution for jets reconstructed from TCCs and topo-clusters, clearly showing the exceptional performances of TCCs over the entire p_T range. This tendency is especially marked at high p_T where the resolution of D_2 using TCCs is improved by a factor of 2 compared to the one using topo-clusters.

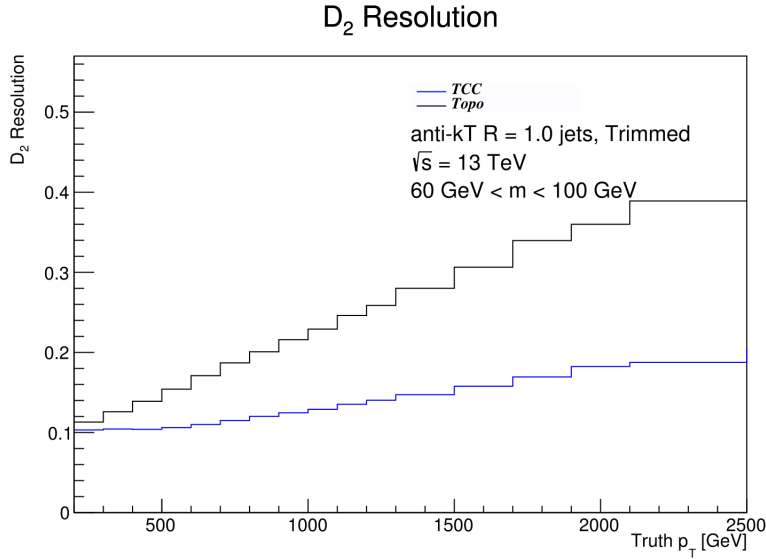


Figure 4.15: Comparison of the D_2 resolution for jets reconstructed using TCCs and topo-clusters in Pythia8 W' samples.

4.5.3 Angularity

Introduction

Angularity is a variable related to the symmetry of the energy distribution inside a jet [43]. It is defined as

$$\tau_a = \frac{1}{M} \sum_i E_i \sin^a \theta_i [1 - \cos \theta_i]^{1-a}$$

with M the mass of the jet, E_i the energy of particle i inside the jet, θ_i its angle to the jet axis and a is a parameter related to the weight given to particles depending on their position. If $a > 0$, more weight is given to particles in the core of the jet and if $a < 0$, more is given to the ones at the edges [44]. The variable is

IR-safe and is used as a discriminator between QCD and boosted bosons because the Angularity distribution displays a bump at small values in its spectrum for boosted vector bosons after mass and D_2 cuts, corresponding to the situation where the two hard constituents are in a symmetric p_T configuration with respect to the jet axis [43]. Figure 4.16 compares the signal and background distributions without mass cut requirements and after a mass cut, then a D_2 cuts have been applied.

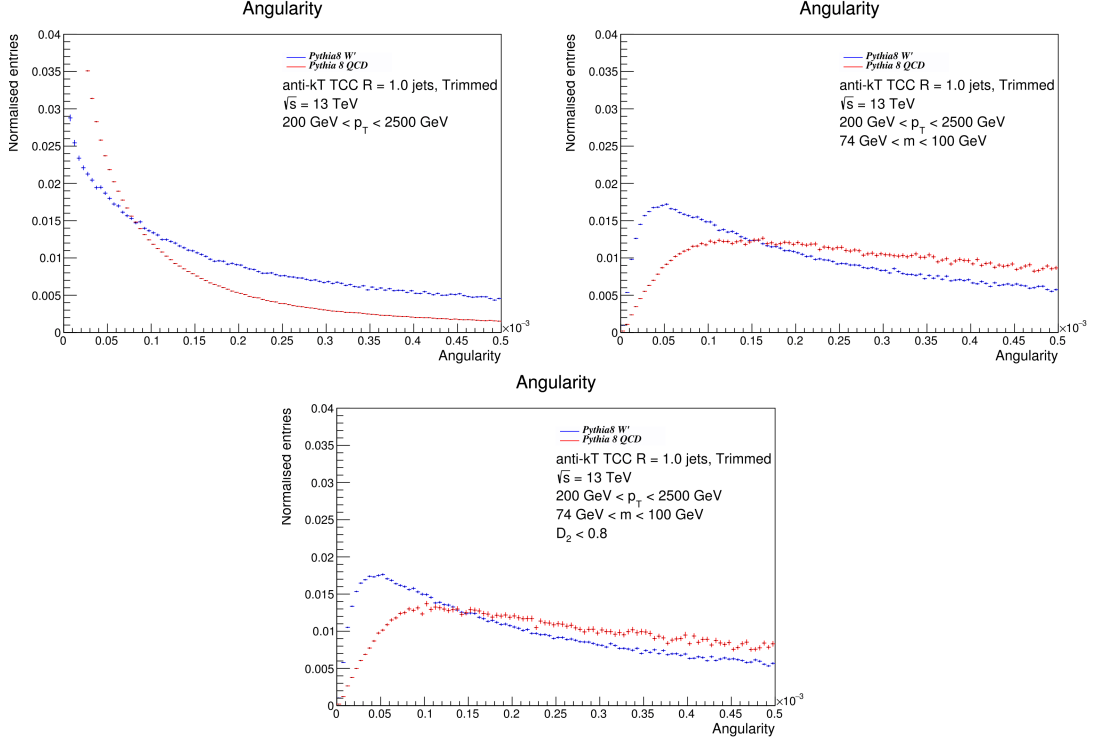


Figure 4.16: Comparison between the Angularity distributions of QCD and W jets, before mass window requirement (top left), after a mass cut of $60 < m < 100$ GeV has been applied (top right) and after mass and D_2 cuts with $D_2 < 1.4$ (bottom).

Applying mass and D_2 cuts shape the Angularity distribution but do not remove all differences between signal and background jets. Angularity is therefore a useful third variable.

Angularity resolution

The performance of TCCs and topo-clusters are compared by computing the resolution of the Angularity derived from TCC reconstructed jets or topo-clusters ones. The definitions of the residual and of the IQR used to obtain the results are the same as the ones used for D_2 resolution calculations and are displayed in Figure 4.17. The resolution of the Angularity derived from TCCs is expected to be slightly worse than the one obtained from topo-clusters. Indeed, despite its name, Angularity mainly uses energy information and has little dependency on angular one as can be seen from its definition. It is confirmed by Figure 4.17 where it can be seen that the resolution of the variable improves with increasing p_T . As p_T resolution improves with increasing p_T and angular resolution gets worse, this further emphasizes the high dependency of the variable on the energy and its low dependency on angles. The

behaviour of the resolution is opposite to the one of mass and D_2 which could give an indication of the usefulness of the Angularity as a substructure variable. Indeed, its resolution improves despite the more collimated configuration of the event and the degrading detector resolutions.

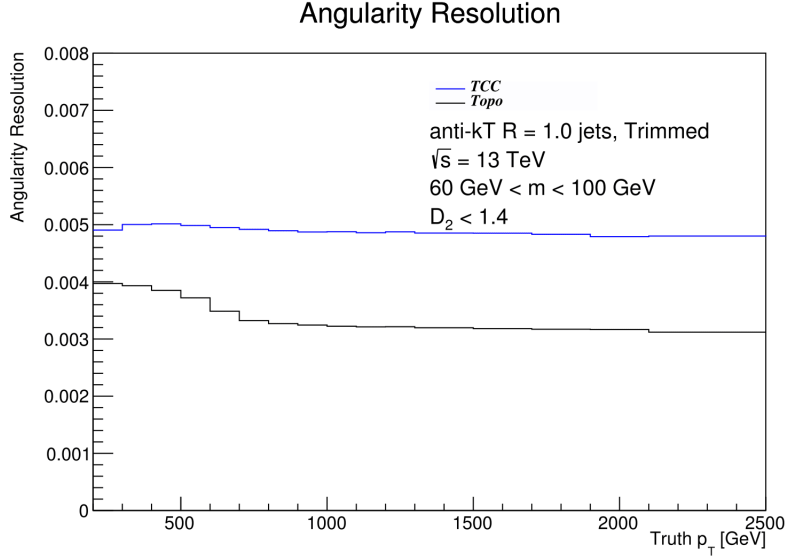


Figure 4.17: Comparison of the Angularity resolution for jets reconstructed using TCCs and topo-clusters in Pythia8 W' samples.

4.5.4 KtDR

Introduction

Jets in ATLAS are reconstructed using the k_T family algorithms, which cluster the inputs to the algorithm based on their distance to each other, 4.2.2. KtDR is the variable entering the distance definition when the jet is reclustered using the k_T algorithm, that is the angular distance ΔR_{ij} between the objects i and j , which corresponds to

$$\Delta R_{ij} = \sqrt{(\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2}$$

KtDR is therefore the distance corresponding to the final clustering step in the pseudorapidity-azimuthal ($\eta - \phi$) space, 4.2.1. Figure 4.18 and Figure 4.19 illustrate the variable differences for the W' and QCD distributions without mass restrictions, after mass cut and after mass and D_2 cuts for low- p_T jets (Figure 4.18) and high- p_T ones (Figure 4.19). The more the energy increases, the more the quarks are collimated and the shorter the distance between the two will be. KtDR will therefore have a particularly large discriminating power at low- p_T . This is shown in the plots of Figure 4.18 and is due to the huge separation of the signal and background distributions, especially after the requirements on mass and D_2 .

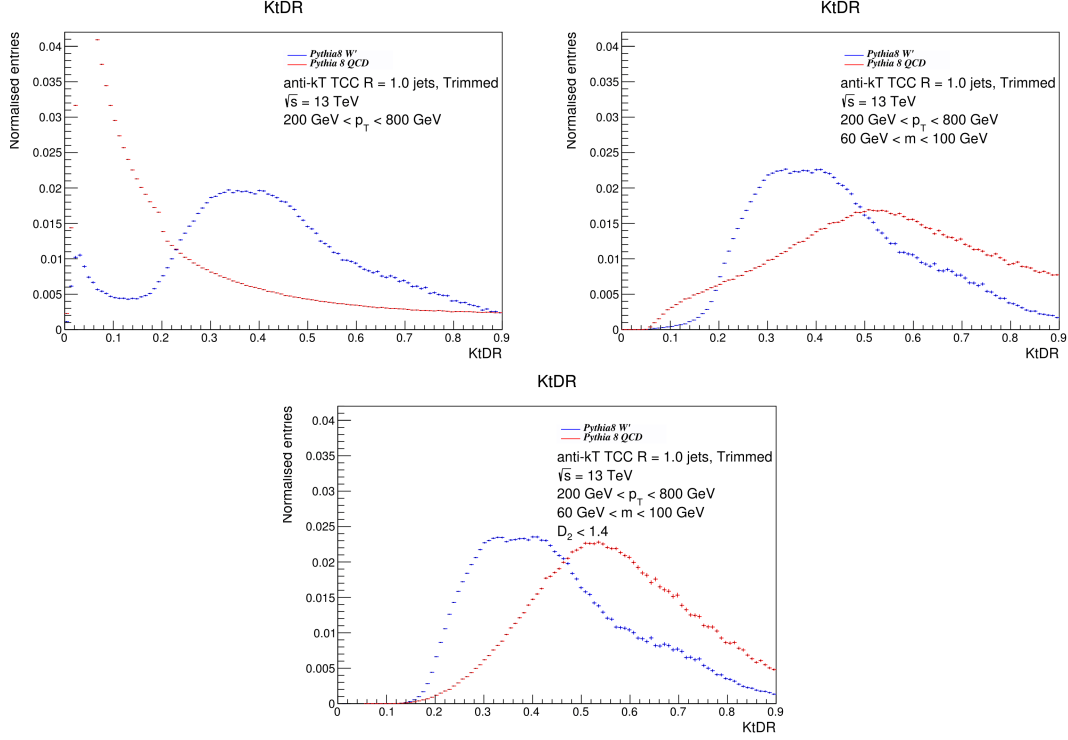


Figure 4.18: Comparison between the KtDR distributions of QCD and W jets with $200 \text{ GeV} < p_T < 800 \text{ GeV}$, before mass window requirement (top left), after a mass cut of $60 < m < 100 \text{ GeV}$ has been applied (top right) and after mass and D_2 cuts with $D_2 < 1.4$ (bottom).

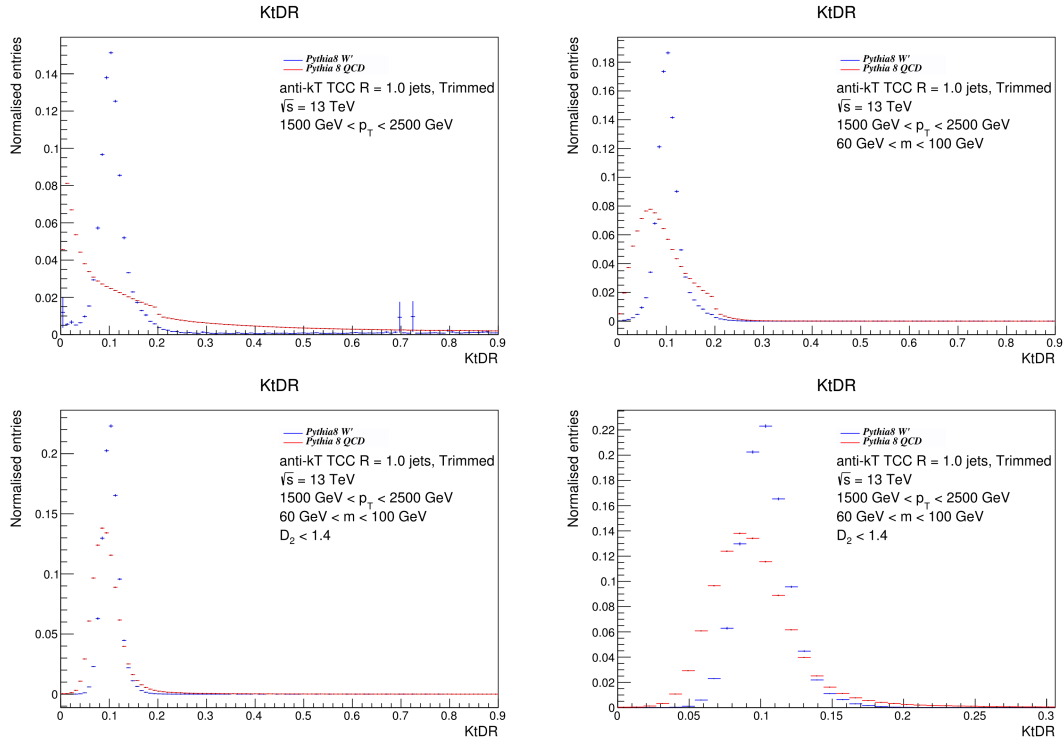


Figure 4.19: Comparison between the KtDR distributions of QCD and W jets with $1500 \text{ GeV} < p_T < 2500 \text{ GeV}$, before mass window requirement (top left), after a mass cut of $60 < m < 100 \text{ GeV}$ has been applied (top right) and after mass and D_2 cuts with $D_2 < 1.4$ (bottom left and right).

KtDR resolution

The comparison of the KtDR resolution obtained by using TCCs and topo-clusters again reveals that TCCs give the best performances thanks to their better angular resolution which is obviously directly linked to the angular distance KtDR. The resolution was computed using the same procedure as for D_2 and Angularity, and the result is illustrated in Figure 4.20. It confirms that TCC-jets significantly outperforms topo-cluster jets over the entire p_T range. Once again, it is important to note the behaviour of the resolution as p_T increases, which is again opposite to the ones of mass and D_2 but similar to the Angularity. This indicates that combining mass, D_2 and KtDR or Angularity could improve the performances of boosted W/Z tagging.

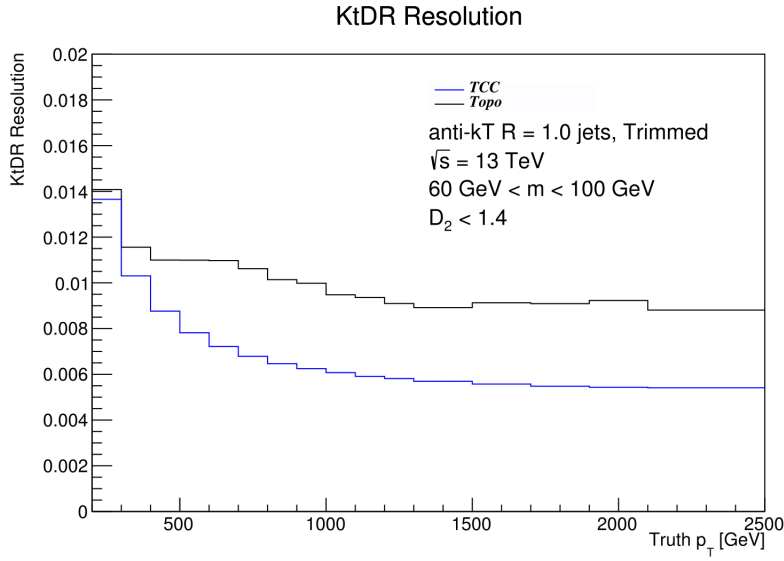


Figure 4.20: Comparison of the KtDR resolution for jets reconstructed using TCCs and topo-clusters for Pythia8 W' samples.

4.5.5 Conclusions

The list of substructure variables presented above is non-exhaustive, and many others exist and are used depending on the particle we want to identify. The aim of this thesis is to build an optimised three variables tagger for boosted bosons identification, using an additional substructure variable to the mass- D_2 tagger already studied [39]. The motivation as to why Angularity and KtDR were chosen is explained in the next chapter, along with the performances comparison between 2-variable and 3-variable taggers. It is also interesting to note that the resolution of most variables presented above seems to motivate the use of TCCs instead of topo-clusters as inputs to the jet reconstruction. This is taken into account in the next chapter as well, as it is expected to improve the tagging efficiency of the 3-variable tagger.

5.1 Motivation

This thesis focuses on diboson production, that is the production of two vector bosons WW , ZZ or WZ decaying from an hypothetical particle W' ($W' \rightarrow WZ$) or G ($G \rightarrow ZZ$, $G \rightarrow WW$). With the high energy scale of $\sqrt{s} = 13$ TeV reached at the LHC, the resulting objects have a large p_T and consequently their decay products ($W' \rightarrow WZ \rightarrow q\bar{q}q\bar{q}$) can only be identified by analysing the substructure of the large- R jets. A simple tagger based on a selection of the jet mass combined with another substructure variable is traditionally used in ATLAS to efficiently tag W bosons. Previous studies [45] have shown that the second optimal variable to be used along with the mass for W taggers is D_2 . However, machine learning (ML) techniques proved that combining these two variables with multiple others lead to a higher discrimination power between signal and background for the resulting multivariate W tagger [46]. Boosted decision trees (BDTs) and deep neural networks (DNNs) have been used to discriminate W -jets from QCD-jets by combining 12 substructure variables and this tagger significantly outperforms two variable taggers for topo-clusters jets [46, 47]. *The idea behind this thesis is: can we get most of the performance gain by using only one additional variable, resulting in a three variable tagger?* Applying simultaneous restrictions on strongly correlated variables can lead to diminishing returns in the tagging efficiency thus imposing a requirement on only one additional variable which is strongly non-correlated to mass and D_2 can lead to similarly powerful tagger. By looking at the linear correlation coefficient matrices of $W \rightarrow qq$ jets (top Figure 5.1) and QCD jets (bottom Figure 5.1), it can be seen that D_2 is substantially uncorrelated with the mass for W jets and anti-correlated with it for QCD jets, which result in a good tagger when combining the two.

Following the same reasoning for other variables with the same correlations behaviours, but this time compared to D_2 instead of the mass, hint that KtDR and Angularity [43] are promising variables satisfying these criteria. To quantify the performances of the taggers, the value of the *significance* is compared for each tagger over the entire p_T range.

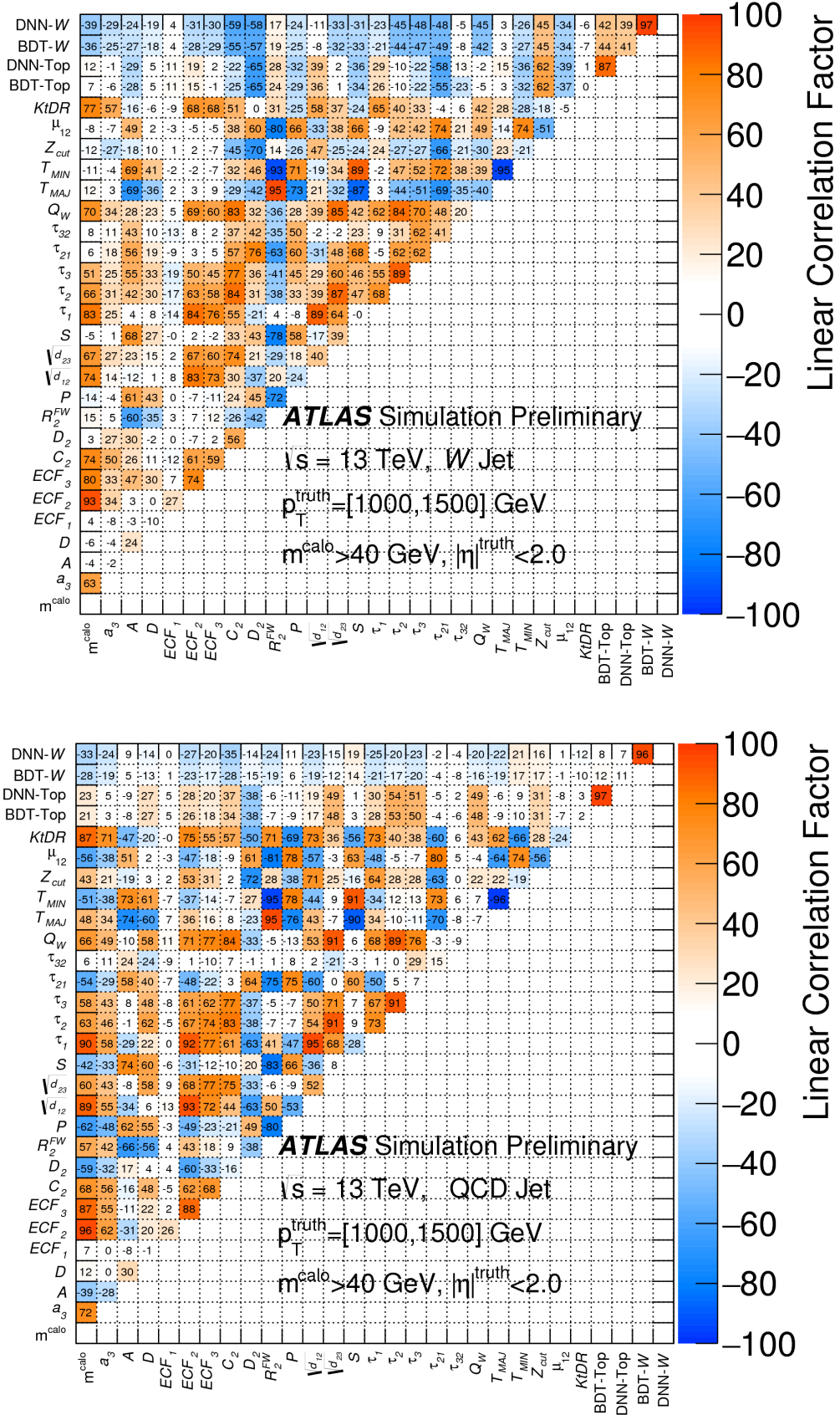


Figure 5.1: Linear correlation coefficient matrices for W jets (top) and QCD jets (bottom) [46]. KtDR and Angularity (the latter referred in the matrices as a_3) are uncorrelated with D_2 for W jets and anti-correlated with it for QCD jets.

This quantity is defined as

$$\text{Significance} = \varsigma = \frac{\epsilon_{sig}}{\frac{3}{2} + \sqrt{\mathcal{E}_{bkg}}}$$

with ϵ_{sig} and \mathcal{E}_{bkg} respectively the *signal efficiency* and the *background rejection* defined as

$$\text{Signal efficiency} = \epsilon_{sig} = \frac{N_{signal}^{selected}}{N_{signal}^{tot}}$$

$$\text{Background rejection} = \mathcal{E}_{bkg} = N_{background}^{selected}$$

Consequently, the best tagger will be the one giving the highest significance per p_T bin, maximally selecting the signal and minimally selecting the background. This definition of the significance is based on the optimization of the sensitivity for searches for new physics described in Reference [48]. The studies mentioned above [46, 47] have been performed using topo-clusters as inputs to the jet reconstruction algorithm. However, TCCs were shown to provide a better resolving power for sub-structure variables, especially at high p_T . ***The additional question asked in this search is: could the performance of the tagger be improved by using jets built from TCCs?*** In this chapter, the traditional two variables tagger is presented and compared with the newly built three variables tagger. For both cases, an additional comparison is performed, first using topo-clusters and then using TCCs. To finish, the overall performances of the taggers are presented in the last section of the chapter.

5.2 Two-variable tagger

5.2.1 Selections on mass and D_2

For the traditional simple tagger, all possible combinations of cuts on both mass and D_2 are considered. The combination resulting in the highest significance gives the values for the mass window selection as well as the optimal cut on the D_2 variable, which together defines the tagger. To do so, the peak value of the mass m_{peak} is found and recursive symmetric cuts $m_{peak} - i < m < m_{peak} + i$ are applied, starting with $i = 1$ GeV and increasing the mass window until the whole distribution has been selected. An example of a two-sided cut on the mass is shown in Figure 5.2.

Once we have a list of mass windows with their respective signal efficiencies and background rejections, a single-sided cut is applied afterwards on D_2 for each of the mass window. Again, the cut is recursive starting from $D_2 = 0$ to $D_2 = 6$, with step size 0.1. The example shown in Figure 5.3 pictures a cut of $D_2 < 0.8$ after the mass window $74 \text{ GeV} < m < 100 \text{ GeV}$ has been applied, corresponding to $i = 13 \text{ GeV}$.

The next step is to compute the significance using the resulting signal efficiencies and background rejections for all possible combinations of mass windows and D_2 cuts. This is summarized in a two dimensional histogram where the significance is displayed as a function of the mass window on the x -axis and the cut on D_2 on the y -axis (Figure 5.4). The highest value of the significance is identified and with it, the corresponding values for the cuts on the mass and D_2 which are defined to be

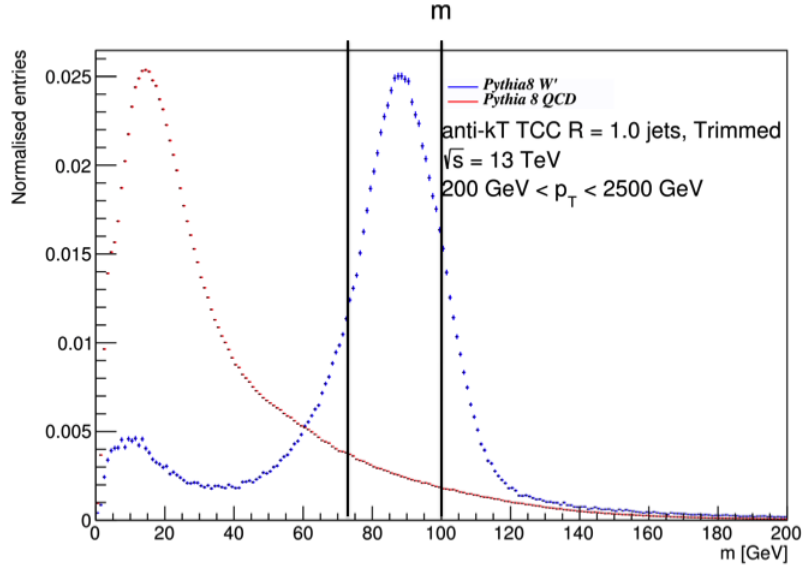


Figure 5.2: Illustration of a two-sided cut on the mass distribution. The example displays a selection on the mass of $74 \text{ GeV} < m < 100 \text{ GeV}$.

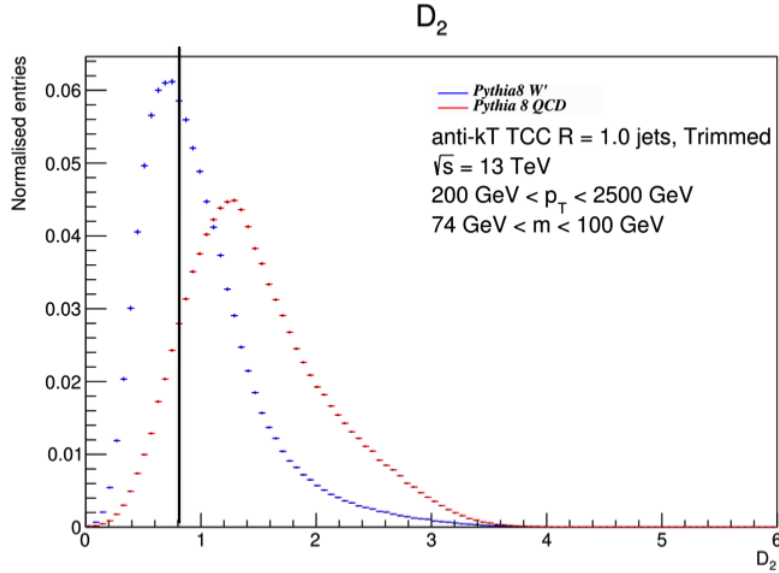


Figure 5.3: Illustration of a single-sided cut on the D_2 distribution. The example displays a cut of $D_2 < 0.8$ after a mass window of $74 \text{ GeV} < m < 100 \text{ GeV}$ has been applied, corresponding to a symmetric cut of $m_{peak} \pm 13 \text{ GeV}$.

the optimal values for the tagger in this p_T range.

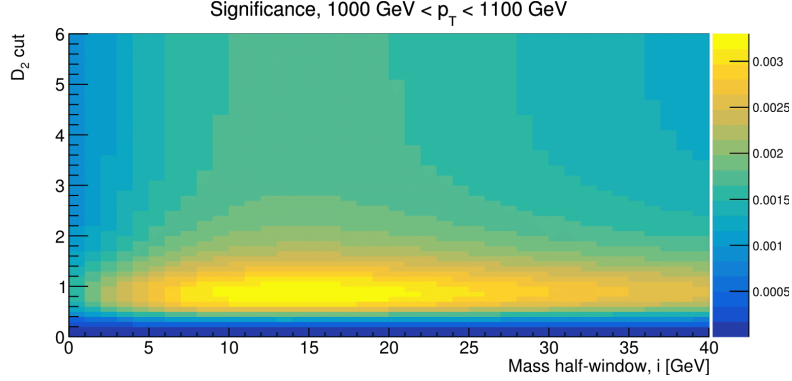


Figure 5.4: Plot of the significance as a function of mass half-window (corresponding to i) and D_2 cuts for $1000 \text{ GeV} < p_T < 1100 \text{ GeV}$ for jets reconstructed using TCCs.

This procedure is done 16 times, once for each p_T window listed in Table 5.1, and therefore 16 values of the significance are extracted. It is important to note that, while D_2 was the best second variable in terms of discriminating power for topo-clusters, this could not have been the case for TCCs. Therefore, other variables were considered but D_2 was found to still be the best second variable when using TCCs as inputs to the jet reconstruction algorithm. The next section compares the performances of the tagger using TCCs and topo-clusters as inputs to the jet reconstruction.

p_T bin	p_T values [GeV]
1	200 - 300
2	300 - 400
3	400 - 500
4	500 - 600
5	600 - 700
6	700 - 800
7	800 - 900
8	900 - 1000
9	1000 - 1100
10	1100 - 1200
11	1200 - 1300
12	1300 - 1500
13	1500 - 1700
14	1700 - 1900
15	1900 - 2100
16	2100 - 2500

Table 5.1: p_T ranges considered.

5.2.2 Comparison using different inputs

The tagging procedure is applied on the Monte Carlo simulated samples. The QCD dijet events are created in 13 p_T slices, going from JZ0W (0 GeV - 20 GeV) to

JZ12W (5300 GeV - infinity). The W' signal samples are produced in 33 samples, by simulating the fully hadronic decay $W' \rightarrow WZ \rightarrow qqqq$ for masses of W' ranging from 500 GeV to 4000 GeV (non-uniform spacing). After the events have been simulated, the jets are reconstructed using both topo-clusters and TCCs as the inputs to the reconstruction. The tagging procedure described above is then applied on the resulting objects and a value of the significance is obtained for each of the p_T bin. Figure 5.5 shows the evolution of the significance as a function of p_T and compares the computed values using TCCs (blue) and topo-clusters (black).

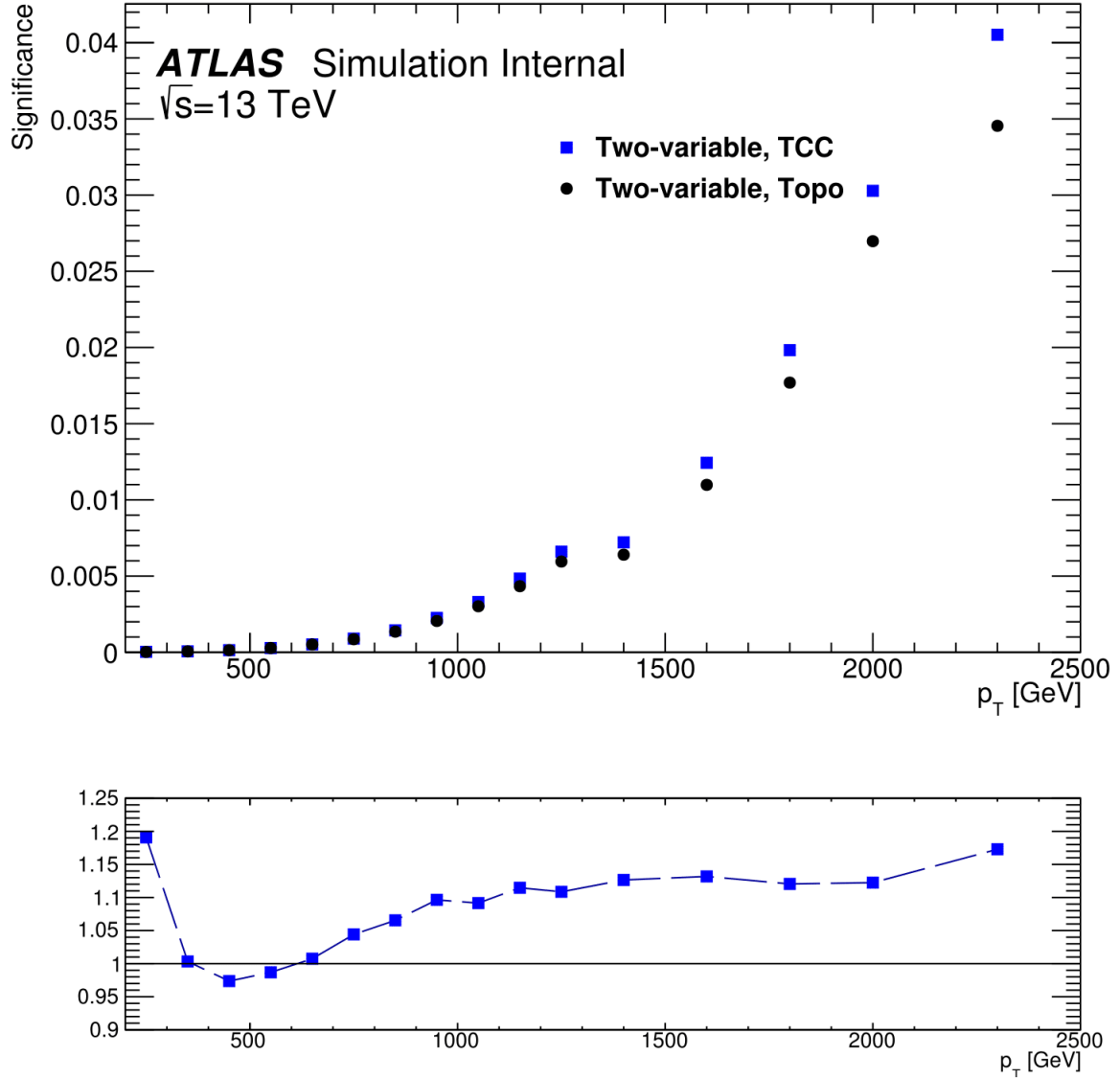


Figure 5.5: (Top) Values of the significance as a function of p_T using the two variables tagger for variables reconstructed using TCCs (blue) and topo-clusters (black). (Bottom) Ratio of significances using TCCs over topo-clusters.

The significance is higher when using TCCs over almost the entire p_T range and particularly excels at high p_T as expected. The performance is quantified in

the bottom plot of Figure 5.5 where the ratio of the significance ς_{TCC} over the significance ς_{topo} is drawn as a function of p_T . As can be seen on the plot, a gain of $\approx 20\%$ is obtained when using TCCs instead of topo-clusters at high p_T values.

5.3 Three-variable tagger

5.3.1 Selections on Angularity and KtDR

The three-variable tagger is obtained by following the same procedure as the one for two variables, selecting mass windows for multiple p_T ranges then for each of the selections, applying a one-sided cut on D_2 and computing the significance for each combination of cuts. In addition to that, a one-sided cut similar to D_2 is applied on the third variable, that is Angularity or KtDR. The resulting signal efficiencies and background rejections are then used to compute the significance for each of the situation. The example in Figure 5.6 displays successive cuts on the three variables: first for a mass window of $74 \text{ GeV} < m < 100 \text{ GeV}$ then for the remaining events a $D_2 < 0.8$ cut is applied and finally a last restriction on the rest of the passed events for the KtDR variable selects events with $\text{KtDR} < 0.46$.

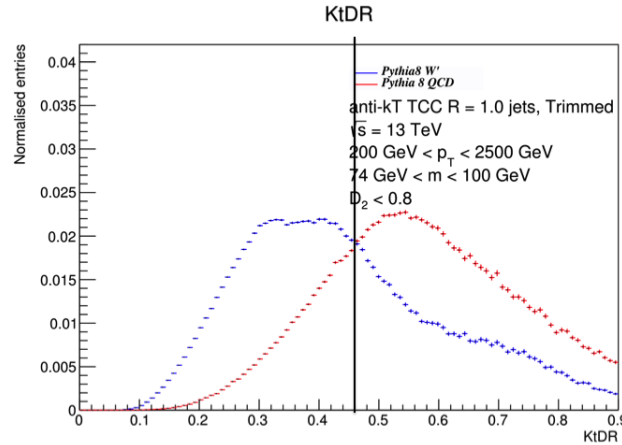


Figure 5.6: Illustration of a single-sided cut on the KtDR distribution. The example displays a cut a restriction of $\text{KtDR} < 0.46$ after a $D_2 < 0.8$ cut and a mass window of $74 \text{ GeV} < m < 100 \text{ GeV}$ have been applied.

This combination of cuts results in a single value of the significance that has to be compared with all the significances computed using the different selections of cuts for this specific p_T bin, the highest one giving the optimal 3 cuts on the 3 variables. All these values are summarized in a 2D histogram where the significance is plotted as a function of D_2 cut on the x -axis and the third variable cut on the y -axis, for a fixed mass window and p_T range. Figure 5.7 shows the histogram for $1000 \text{ GeV} < p_T < 1100 \text{ GeV}$ for a mass window of $\pm 13 \text{ GeV}$ around m_{peak} for KtDR (top) and Angularity (bottom).

The comparison of the significances resulting from TCCs and topo-clusters jets is performed in the next section.

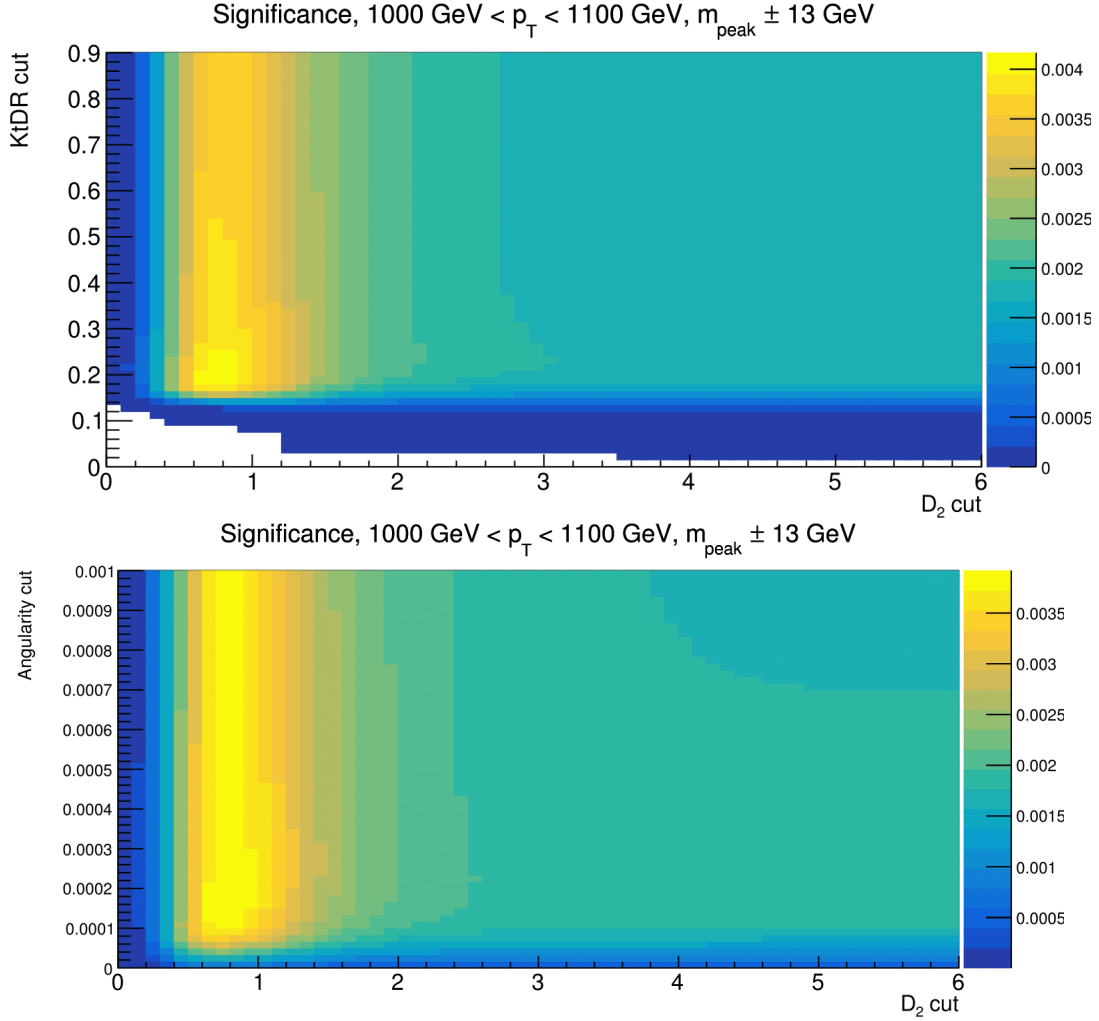


Figure 5.7: Plot of the significance as a function of D_2 and KtDR cuts (top) and D_2 and Angularity cuts (bottom) for p_T values in $1000 \text{ GeV} < p_T < 1100 \text{ GeV}$ with a mass window of $m_{\text{peak}} \pm 13 \text{ GeV}$, for jets reconstructed using TCCs.

5.3.2 Comparison using different inputs

The highest values of the significance for each of the p_T window are plotted in Figure 5.8, comparing the values obtained using TCCs or topo-clusters as inputs to the jet algorithm. Once again, TCCs outperform topo-clusters over almost the entire p_T range and the gain is shown in bottom Figure 5.8, clearly displaying an approximate 30% gain of TCC jets over topo-cluster ones for both variables, especially at high p_T .

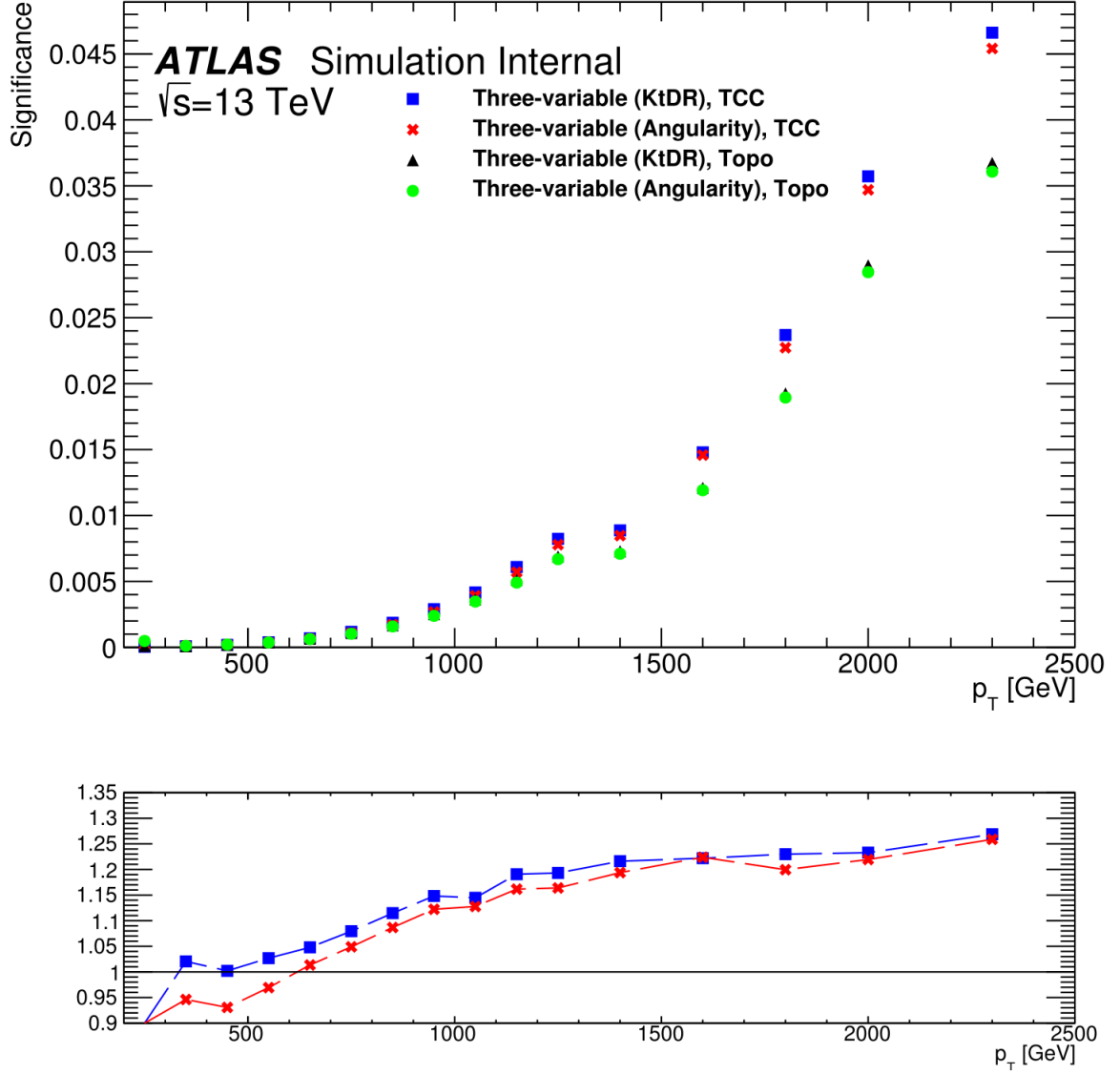


Figure 5.8: (Top) Values of the significance as a function of p_T using KtDR with TCCs (blue) or topo-clusters (black) and using Angularity with TCCs (red) or topo-clusters (green). (Bottom) Ratio of significances using TCCs over topo-clusters for KtDR (blue) and Angularity (red).

5.4 Summary

5.4.1 Significance gain using a third variable

Track-CaloClusters have been developed to increase jet substructure performance as it improves the angular resolution of jet substructure variables, especially at high p_T . Application of the TCC methodology results in values of the significance approximately 20% higher than in the case in which the variable is reconstructed using topo-clusters. This is confirmed both for the two-variable tagger and the three-variable tagger, as described in sections 5.2.2 and 5.3.2. This section focuses on the significance gain when using a third variable, in the case when jets are reconstructed using topo-clusters (top plot of Figure 5.9) and TCCs (bottom plot of Figure 5.9).

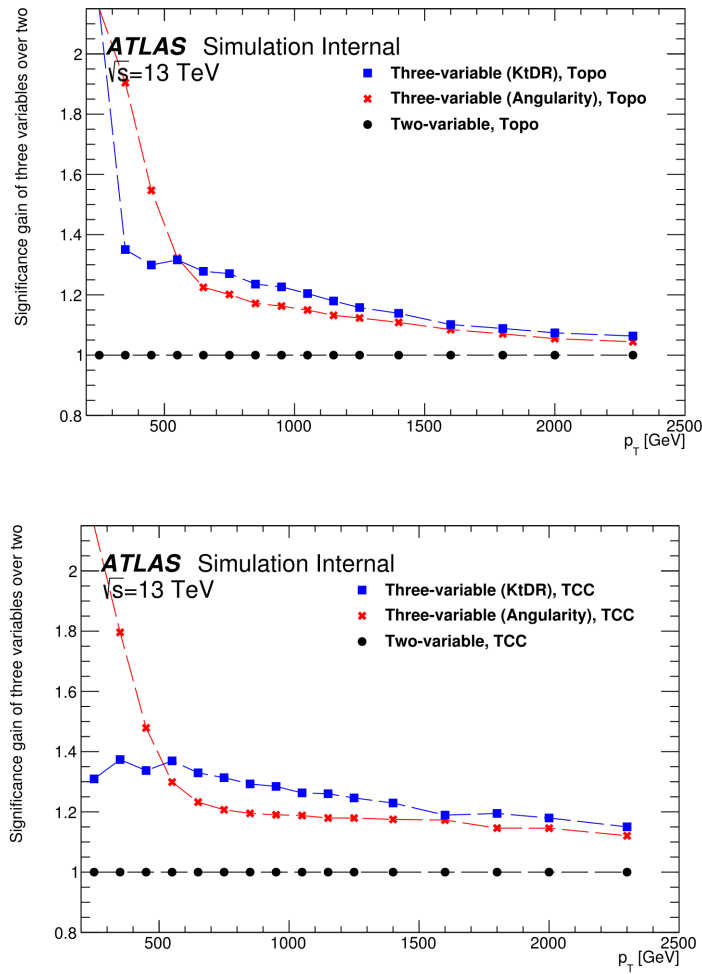


Figure 5.9: Ratio of the significances computed using three variables over two variables for jets reconstructed using topo-clusters (top) and TCCs (bottom).

The use of a third variable for jets reconstructed using TCCs clearly indicates a gain of roughly 20% in the region of interest, that is at high- p_T . As expected, the gain is lower when using topo-clusters as inputs to the jet reconstruction algorithm, with an increase of significance of only 10% at high- p_T .

5.4.2 Significance gain using TCCs and a third variable

The goal of this thesis is to build a new tagger which uses the TCC technology and combines it with the use of a third variable to further improve the discrimination between signal and QCD background. The overall results are displayed in Figure 5.10. It is interesting to start with the Reference (black) which corresponds to the traditional two-variable tagger and to draw comparisons from this point. It can be seen that the use of a third variable without applying the TCC procedure (magenta), results in much larger values of the significances at low- p_T but that these decrease with increasing p_T . This behaviour is opposite to the one when only applying the TCC methodology without adding a third variable (green), which results in increasing values of the significance when increasing p_T . Attention should now be drawn to the central goal of this thesis, that is the combination of these two methods in the building of a new tagger. It is important to notice that it is only when combining TCCs with a third variable that such gains on the significance can be obtained over the entire p_T range. In the $p_T > 500$ GeV region, not only a gain of up to 20% is obtained when applying the TCC procedure (green), but using a third variable leads to an additional roughly 20% gain (blue for KtDR, red for Angularity). This means that the new three-variable tagger results in an overall gain up to 40% in discrimination significance at high p_T when compared to the traditional tagger currently used in ATLAS.

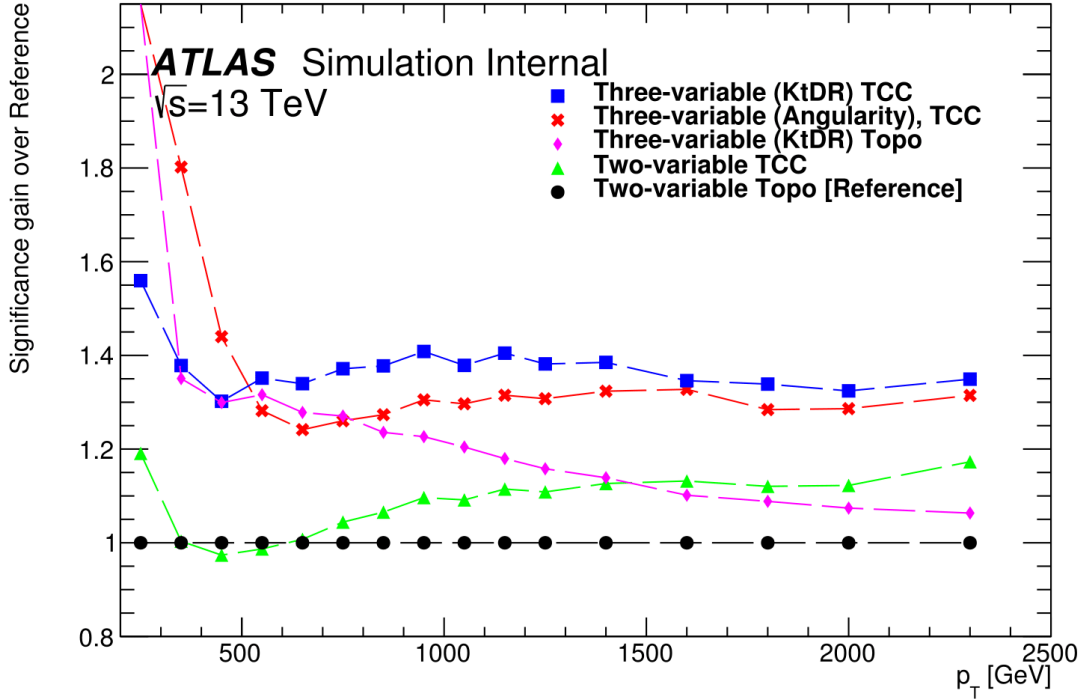


Figure 5.10: Ratio of the significances computed using TCCs over topo-clusters as a function of p_T for the two-variable tagger (green) and using KtDR (blue) and Angularity (red) as the third variable for the three-variable tagger. The two-variable topocluster significance is the Reference.

CHAPTER 6

CONCLUSION

Vector bosons produced at the LHC are often highly boosted due to their transverse momentum being much larger than their mass. Their decay products are therefore collimated and jet substructure techniques are required for the identification of these vector bosons. A requirement on the combined jet mass and a single-sided cut on the D_2 substructure variable are currently used in ATLAS for the discrimination between W/Z -jets and QCD-jets.

The aim of this thesis is to improve the significance with which we discriminate between the two types of jets by using a new tagger. This tagger is built by imposing a requirement on the jet mass and by applying single-sided cuts on D_2 but also on another substructure variable, either KtDR or the Angularity. In addition to that, the tagger no longer uses jets built from topo-clusters but rather the jets are reconstructed using Track-CaloClusters (TCCs). These objects exploit both the excellent energy resolution of the calorimeters and the better spatial resolution of the tracker at high p_T .

The tagger making use of these objects is seen to perform better than the one using substructure variables reconstructed from topo-clusters. In this thesis, comparisons of the values of the significance obtained from the traditional two-variable tagger (using TCCs and topo-clusters) and the one obtained from the new three-variable tagger (using TCCs and topo-clusters) are performed. It results in a significance gain of roughly 20% when using the three-variable tagger which is further increased by approximately 20% when using TCCs as inputs to the jet reconstruction algorithm, corresponding to an overall gain of roughly 40% for the newly built three-variable tagger. As the diboson analysis uses a double-tag, corresponding to the two vector bosons, it is roughly an improvement in the analysis sensitivity by a factor of 2.

This new tagger should therefore increase the efficiency of diboson identification and, consequently, improve the sensitivity for searches for new physics using the ATLAS detector.

ACKNOWLEDGEMENTS

The writing of this thesis would have never been possible without the support from my friends, family and professors and I would like to thank them for playing a huge part in it.

First of all, I would like to thank Prof. Giuseppe Iacobucci who gave me the opportunity to work in the high-energy physics field and hence made me discover the most exciting working-environment I could have ever imagined. This last year of study made me realize that this was exactly what I wanted to do for the coming years; contribute to the wonderful world of scientific research in projects that have a positive impact on the future of humanity, distant or near.

I would also like to thank Noemi and Juno for giving their time in answering all the stupid questions I've asked, with gentleness and sympathy. Thank you to my amazing supervisors, Roland for welcoming me in the group and always be there for additional explanations and Steven without whom this thesis would never be. I cannot thank Steven enough for his advices and lessons, almost everything I've learnt on jet physics and related computing is owed to him. Thank you for your availability, kindness and your unlimited amount of patience in sharing your knowledge.

Very special thanks to my mentor and best friend, Sofia, who was the first person I met in Geneva and has always been there since to guide me. Thank you for being there for a good beer after work, exciting ski trips to Courmayeur and countless amazing dinners at your place. Thank you to Stephanie, Rebecka and Romane for sharing my everyday life, cheering me up in the downs and celebrating with me in the ups. Thank you for your love and happiness, you've made these two years in Geneva the best of my life (yet).

Special credits to my parents, who have always been very supportive in my studies even though they have no idea what a quark is. Thank you for giving me the opportunity to live my dream in Geneva and for being the best parents one could dream of. Thank you to my brother Hugo and my sister Juliette for sharing my interests in countless domains and for making family trips unforgettable, I love you both unconditionally.

Last thanks to Geraldine, Megan, Helene, Juliette, Camille, Alice, Justine, Yannick, Antoine, Heather and Chloe for being precious friends and making my life amazing.

Thank you, to all of you.

Bibliography

- [1] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.
- [2] *The Standard Model*, in https://en.wikipedia.org/wiki/Standard_Model
- [3] S. L. Glashow, *Towards a unified theory: Threads in a tapestry*, 1980, **Rev. Mod. Phys.** **52**, 539.
- [4] A. Salam, *Gauge unification of fundamental forces*, 1980, **Rev. Mod. Phys.** **52**, 525.
- [5] S. Weinberg, *Conceptual foundations of the unified theory of weak and electromagnetic interactions*, 1980, **Rev. Mod. Phys.** **52**, 515.
- [6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, 1964, <https://doi.org/10.1103/PhysRevLett.13.321>.
- [7] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, 1964, **10.1103/PhysRevLett.13.508**.
- [8] Private Discussion, Steven Schramm
- [9] Super-Kamiokande Collaboration, Y. Fukada et al. *Evidence for Oscillation of Atmospheric Neutrinos*, 1998, **10.1103/PhysRevLett.81.1562**.
- [10] *ALPHA*, in <https://home.cern/about/experiments/alpha>
- [11] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, 2005, **arXiv:hep-ph/0404175v2**.
- [12] H. Georgi and S. L. Glashow, *Unity of All Elementary Particle Forces*, 1974, **10.1103/PhysRevLett.32.438**.
- [13] H. Fritz and P. Minkowski, *Unified Interactions of Leptons and Hadrons*, 1975, **10.1016/0003-4916(75)90211-0**.
- [14] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, 1999, **arXiv:hep-ph/9905221**.

- [15] J. R. Andersen et al. *Discovering technicolour*, 2011, **arXiv:1104.1255**.
- [16] ATLAS Collaboration, *Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, 2015, **arXiv:1506.00962**.
- [17] ATLAS Collaboration, *Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, 2017, **arXiv:1708.04445**.
- [18] *Trigger and Data acquisition system*, in <https://atlas.cern/discover/detector/trigger-daq>
- [19] T. Sjostrand and Stephen Mrenna and Peter Skands, *A Brief Introduction to PYTHIA 8.1*, 2008, **arXiv:0710.3820**.
- [20] O. S. Brunning et al., *LHC Design Report. 1. The LHC Main Ring*, 2004, **10.5170/CERN-2004-003-V-1**.
- [21] O. S. Brunning et al., *LHC Design Report. 2. The LHC infrastructure and general services*, 2004, **10.5170/CERN-2004-003-V-2**.
- [22] M. Benedikt et al., *LHC Design Report. 3. The LHC injector chain*, 2004, **10.5170/CERN-2004-003-V-3**.
- [23] C. Lefevre, *LHC: the guide*, 2008, <https://cds.cern.ch/record/1092437>.
- [24] *LHC Commissioning*, <http://lhc-commissioning.web.cern.ch/lhc-commissioning/>.
- [25] *High Luminosity Large Hadron Collider*, <http://hilumilhc.web.cern.ch/about/hl-lhc-project>.
- [26] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, 2008, **10.1088/1748-0221/3/08/S08003**.
- [27] C. W. Fabjan and F. Gianotti, *Calorimetry for Particle Physics*, 2003, **CERN-EP/2003-075**.
- [28] A. Martyniuk, *A search for diboson resonances at ATLAS using boson-tagged jets*, 2016 <http://www.ep.ph.bham.ac.uk/general/seminars/slides/Alex-Martyniuk-2016.pdf>.
- [29] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, 2015, **arXiv:1603.02934**.
- [30] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s}=7$ TeV with the ATLAS detector*, 2015, **arXiv:1406.0076**.
- [31] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV*, 2013, **arXiv:1112.6426**.

- [32] ATLAS Collaboration, *Performance of jet substructure techniques for large- R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, 2013, **arXiv:1306.4945**.
- [33] ATLAS Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, 2017, **ATL-PHYS-PUB-2017-015**.
- [34] G.P. Salam, *Towards Jetography*, 2010, **arXiv:0906.1833**.
- [35] J.E. Huth et al., *Toward a standardization of jet definitions*, 1990, **FNAL-C-90-249-E**.
- [36] M. Cacciari and G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, 2008, **arXiv:0802.1189**.
- [37] D. Krohn and J. Thaler and L-T. Wang, *Jet Trimming*, 2010, **arXiv:0912.1342**.
- [38] S. Schramm, *ATLAS Jet Reconstruction, Calibration, and Tagging*, 2013, **ATL-PHYS-SLIDE-2017-781**.
- [39] ATLAS Collaboration *Performance of Top Quark and W Boson Tagging in Run 2 with ATLAS*, 2017, **ATLAS-CONF-2017-064**.
- [40] ATLAS Collaboration *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, 2016, **ATLAS-CONF-2016-035**.
- [41] A. J. Larkoski and I. Moult and D. Neill, *Power Counting to Better Jet Observables*, 2014, **arXiv:1409.6298**.
- [42] A. J. Larkoski and G. P. Salam, J. Thaler, *Energy Correlation Functions for Jet Substructure*, 2013, **arXiv:1305.0007**.
- [43] ATLAS Collaboration, *ATLAS Measurements of the Properties of Jets for Boosted Particle Searches*, 2012, **arXiv:1206.5369**.
- [44] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, I. Sung and J. Virzi, *Substructure of high- p_T Jets at the LHC*, 2009, **arXiv:0807.0234**.
- [45] ATLAS Collaboration, *Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s}=13$ TeV Monte Carlo Simulations for ATLAS*, 2015, **ATL-PHYS-PUB-2015-033**.
- [46] ATLAS Collaboration, *Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s} = 13$ TeV*, 2017, **ATL-PHYS-PUB-2017-004**.
- [47] ATLAS Collaboration, *Performance of Top Quark and W Boson Tagging in Run 2 with ATLAS*, 2017, **ATLAS-CONF-2017-064**.
- [48] G. Punzi, *Sensitivity of searches for new signals and its optimization*, 2003, **arXiv:physics/0308063**.

- [49] F. Guescini, *Search for Resonances in the Dijet Mass Spectrum with the ATLAS Experiment at the LHC*, 2016, **CERN-THESIS-2016-235.5**.
- [50] C. M. Delitzsch, *Search for Diboson Resonances in the Fully Hadronic Final State Using Jet Substructure Techniques in 8 and 13 TeV Proton-Proton Collisions with the ATLAS Detector*, 2016, **CERN-THESIS-2016-236**.
- [51] A. Picazio, *Search for heavy di-boson resonances decaying to boosted hadronic final states in proton-proton collision at $\sqrt{s}=8$ TeV with the ATLAS detector*, 2016, in http://dpnc.unige.ch/THESES/THESE_PICAZIO.pdf.
- [52] S. Adorni, *Improving jet substructure performance in ATLAS by unifying tracking and calorimeter information*, 2017, **SAPIENZA Master thesis**.