Search for emerging jets with the ATLAS detector and reinterpretation of the LHC results

Thomas Wojtkowski

To cite this version:

Thomas Wojtkowski. Search for emerging jets with the ATLAS detector and reinterpretation of the LHC results. Physics [physics]. Université Grenoble Alpes [2020-..], 2025. English. ⟨NNT : 2025GRALY029⟩. ⟨tel-05372672⟩

HAL Id: tel-05372672
https://theses.hal.science/tel-05372672v1

Submitted on 19 Nov 2025

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : PHYS - Physique
Spécialité : Physique Subatomique et Astroparticules
Unité de recherche : Laboratoire de Physique Subatomique et Cosmologie

Recherche de jets émergents avec le détecteur ATLAS et réinterprétation des résultats du LHC

Search for emerging jets with the ATLAS detector and reinterpretation of the LHC results

Présentée par :

Thomas WOJTKOWSKI

Direction de thèse :

Marie-Hélène GENEST — Directrice de thèse
DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES

Pierre-Antoine DELSART — Co-encadrant de thèse
CHARGE DE RECHERCHE, CNRS DELEGATION ALPES

Rapporteurs :

Jérémy ANDREA
DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALSACE
Julien DONINI
PROFESSEUR DES UNIVERSITES, UNIVERSITE CLERMONT AUVERGNE

Thèse soutenue publiquement le **9 octobre 2025**, devant le jury composé de :

Jérémy ANDREA, DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALSACE	Rapporteur
Julien DONINI, PROFESSEUR DES UNIVERSITES, UNIVERSITE CLERMONT AUVERGNE	Rapporteur
Corinne GOY, DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES	Présidente
Stéphanie BEAUCERON, CHARGEE DE RECHERCHE, CNRS DELEGATION PARIS CENTRE	Examinatrice
Fabrice BALLI, INGENIEUR DOCTEUR, CEA SACLAY	Examinateur
Marine KUNA, MAITRESSE DE CONFERENCES, UNIVERSITE GRENOBLE ALPES	Examinatrice

Contents

Introduction

The Standard Model of particle physics is a theory that allows to describe the elementary particles and the existing fundamental interactions. These particles all have been detected and many of their properties measured, the last one being the Higgs boson discovered in 2012 at the Large Hadron Collider by the ATLAS and CMS experiments.

However, the Standard Model remains an incomplete theory since it is unable to explain several physical observations such as the nature of dark matter, an invisible form of matter evinced through its gravitational effects. The nature of dark matter remains unknown, even if a large number of hypotheses has been explored by physicists. A possibility would be to consider that dark matter is made of new invisible particles that interact weakly with the visible ones, and that could be produced at hadron colliders.

This thesis focuses on theoretical models arising from a dark sector accessible through a new mediator, containing new particles and able to provide dark matter candidates. More precisely, the experimental signature that has been considered corresponds to jets of particles associated with displaced vertices, generally referred to as «emerging jets».

In this thesis, chapter 1 briefly describe the Standard Model and focus especially on the strong interaction, since it is the interaction that is the more related to these studies. Chapter 2 presents the several evidences for the presence of dark matter in the universe, in addition to the main hypotheses to explain its nature. Chapter 3 presents the Large Hadron Collider and the ATLAS detector, while Chapter 4 describes the method employed in ATLAS to reconstruct the physical objects used in this thesis, along with the studies I have done on hadronic jets. Then, chapter 5 presents my contribution to the search for emerging jets with the ATLAS detector using the LHC Run-3 data that led to an article that has been submitted to the journal «Reports on Progress in Physics». Finally, chapter 6 presents the concept of reinterpreting existing LHC analyses, in addition to my implementation of analyses in the reinterpretation framework «MadAnalysis» and the first studies performed to allow the reinterpretation of the emerging jet analysis.

Chapter 1

The Standard Model of particle physics

Contents

1.1 The Standard Model of particle physics

The Standard Model of particle physics is a theory describing three of the four fundamental forces which are the electromagnetic, the weak and the strong interactions, by classifying the fundamental particles that constitute matter and quantifying their interactions. It provides predictions that are in excellent agreement with experimental measurements, making the Standard Model one of the most successful theory of physics.

The Standard Model is characterized as a gauge quantum field theory that is based on local symmetries, and that considers each particle as an excitation of a quantum field. It is defined by the gauge group $SU(3) \otimes SU(2) \otimes U(1)$, where $SU(3)$ is the gauge group of the strong interaction also called Quantum Chromodynamics (QCD) and $SU(2) \otimes U(1)$ is the gauge group of the electro-weak interaction. Through the Higgs mechanism [1], the electro-weak gauge symmetry is spontaneously broken, separating the electromagnetic and weak interactions and explaining how the massive particles can acquire mass.

The Standard Model particle contents is presented in Figure 1.1 [2]. It is made of two types of particles : fermions of spin 1/2 that compose matter and gauge bosons of spin 1 that are the mediators of the fundamental interactions. It also contains the Higgs boson H, a spin 0 boson originating from the Higgs mechanism.

The electromagnetic interaction that only occurs between electrically charged particles is mediated by the photon γ which is a massless and electrically neutral particle. The strong interaction that occurs between particles that carry a color charge is mediated by eight different massless gluons

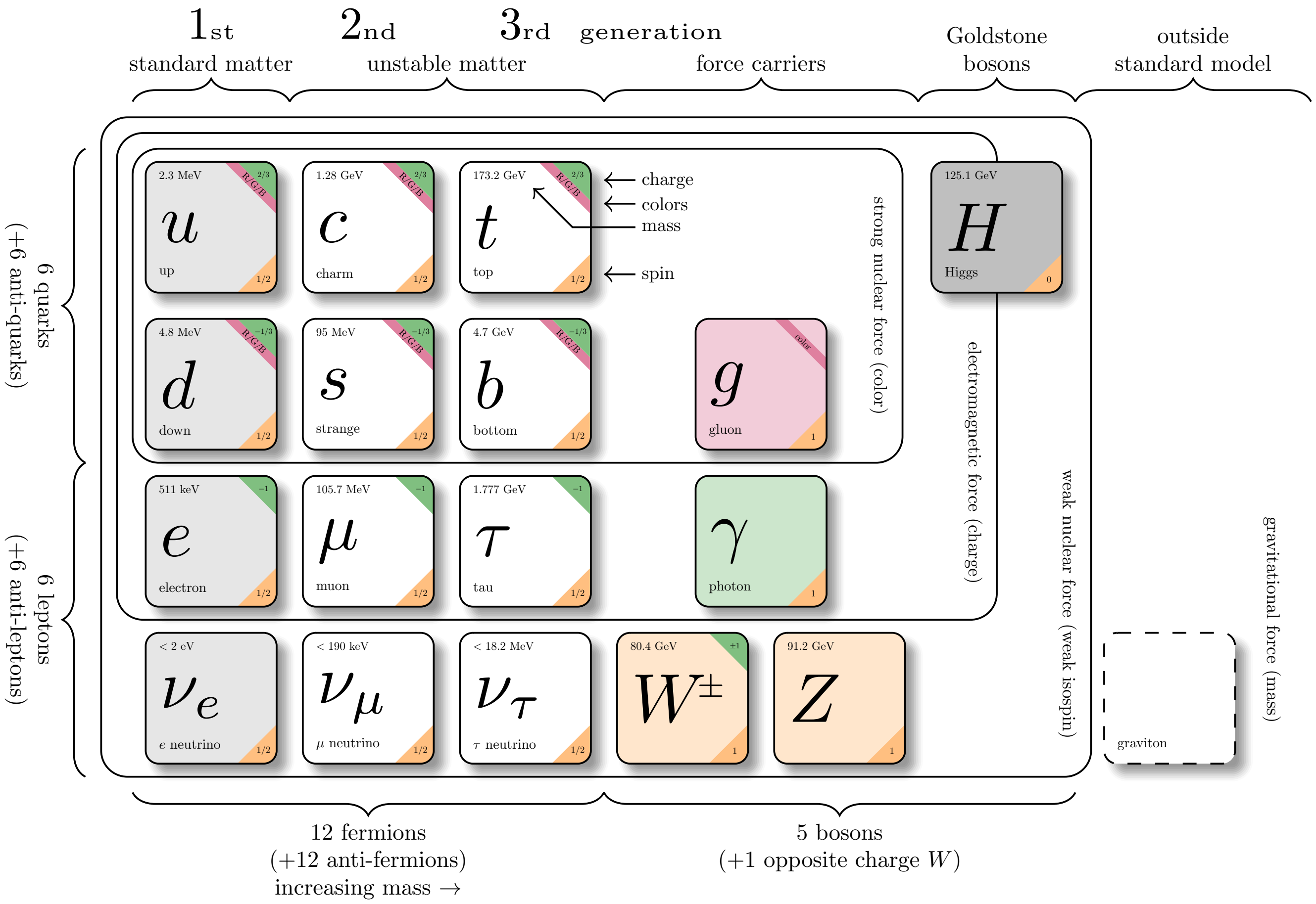


Figure 1.1: Diagram of the Standard Model of particle physics presenting the known fundamental particles [2].

g that have a color - anticolor charge, allowing interactions between gluons. The weak interaction that affects all the fermions is mediated by the massive bosons W^+, W^- and Z^0.

The fermions are divided into quarks and leptons, which are both classified into three generations distinguished by their hierarchy in mass.

The first quark generation is made of the up and down, the second one of the charm and strange, and the third one of the top and bottom quarks. The up, charm and top carry an electric charge of $+\frac{2}{3}$, while the down, strange and bottom have an electric charge of $-\frac{1}{3}$. The six quarks also carry a color charge, allowing them to interact with gluons, additionally to photons, $W^\pm$ and Z bosons. Because they are the lightest ones, only the quarks from the first generation are stable and form protons and neutrons.

The leptons are composed of particles that carry an electric charge of -1 which are the electron e, the muon μ and the tau τ, and of neutral particles ν_e, ν_μ and ν_τ respectively named electron, muon and tau neutrinos. The first lepton generation is made of e and ν_e, the second one of μ and ν_μ, and the third one of τ and ν_τ. Both charged leptons and neutrinos can interact with the $W^\pm$ and Z bosons through the weak interaction but only the charged leptons can interact with photons. As they do not carry color, no leptons can interact with gluons. Because it is lighter than the muon and the tau, the electron is the only charged lepton that is stable and allows to form atoms together with protons and neutrons, constituting the baryonic matter.

For each of these fermions, a corresponding antiparticle also exists with the same mass but with opposite quantum numbers (like the electric charge). For example, the antiparticule of the electron is named the positron e^+ and has a charge equal to $+1$.

All of these particles have been detected, the latest being the top quark in 1995 ([3], [4]) and the tau neutrino in 2000 [5] at Fermilab, and the Higgs boson in 2012 ([6], [7]) at the Large Hadron Collider, corresponding respectively to the most massive fermion, one of the weakest interacting fundamental particle and the most massive boson in the Standard Model.

Despite its success, the Standard Model presents some limitations and is unable to explain some physical observations. Indeed, among others, it does not describe the gravitational interaction, does not explain the observation that the neutrinos are massive, cannot explain the matter - anti-matter asymmetry observed in the universe, and does not provide any candidate for the dark matter. In particular, the dark matter problem will be more detailed in section 2.

1.2 Quantum Chromodynamics

In this section, the strong interaction described by the gauge quantum field theory named Quantum Chromodynamics (QCD) will be more detailed as it is highly related to the studies presented in this document. As explained in the previous section, QCD is described by the $SU(3)$ gauge group. It describes the interactions between the quarks with the gluons as mediators. Each quark can carry one of three possible color charges (Red, Green and Blue) while each of the eight different gluons carry a color - anticolor charge.

QCD presents some notable properties. Indeed, gluons are color-charged particles so gluon self-interactions are possible as illustrated in Figure 1.2 [8] with 3-gluon and 4-gluon vertices, while photon self-interactions through the electromagnetic interaction are forbidden because the photon is a neutral particle.

The QCD coupling constant α_S is actually not constant but vary with the energy scale of the interaction. It can be written as :

$$\alpha_S(Q^2) = \frac{1}{\beta \ln(\frac{Q^2}{\Lambda^2})} \tag{1.1}$$

where Q^2 is the square of the four-vector transferred during the interaction, β is a known positive

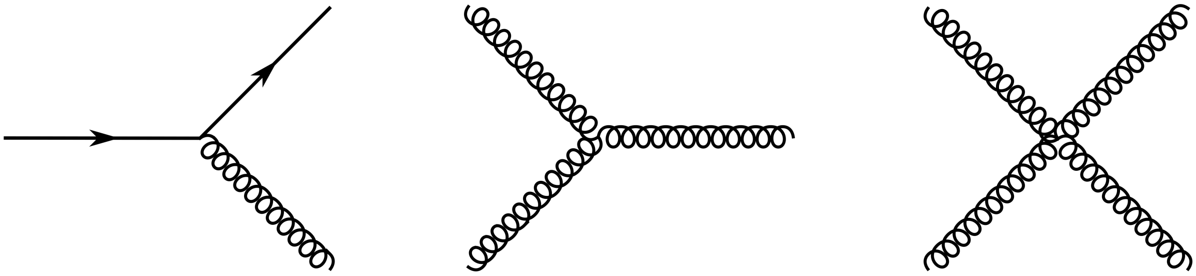


Figure 1.2: Feynman diagrams representing quark-gluon, 3-gluon and 4-gluon QCD vertices [8].

constant and Λ is called the QCD energy scale and is about a few hundred of MeV. As a consequence, α_S will decrease if Q^2 increases, signifying that α_S is smaller at high energy than at low energy (or equivalently, α_S is smaller at short distance than at large distance) while the electromagnetic coupling constant presents an opposite behavior. This property of the QCD is named the asymptotic freedom and allows the free propagation of quarks and gluons in high energy processes ($Q^2 \gg \Lambda$).

On the other hand, in low energy processes ($Q^2 \leq \Lambda$), quarks and gluons cannot be isolated because the interaction between them becomes much stronger, leading to the confinement of color-charged particles into composite states named hadrons that have a color charge equal to zero. This property of the QCD is called the color confinement. The hadrons are classified either as baryons (such as the protons, the neutrons...) if they are composed of three quarks, or as mesons if they are composed of a quark - anti-quark pair (such as the pions, the rhos, the kaons...)

The evolution of α_S as a function of Q obtained from different measurements is shown in Figure 1.3 [9]. For exemple, the α_S measured value at an energy corresponding to the mass of the Z boson is equal to $\alpha_S(M_Z^2) = 0.1179 \pm 0.0010$ [9]. In high energy hadron collisions, a process can be separated between a hard perturbative contribution and a softer non-perturbative one that are treated differently in the computations.

1.3 The hadron colliders

At the TeV energy scale relevant to the Large Hadron Collider (LHC) proton-proton collisions, it is not the protons but the quarks and gluons that compose them, denoted as partons, that interact. The partons each carry a fraction of the energy of the proton they compose. In a proton-proton collision, an interaction between two energetic partons can occur and produce the emission of an energetic parton (or a massive gauge boson), carrying a large momentum in the transverse plane of the collision. This interaction is named hard scattering process.

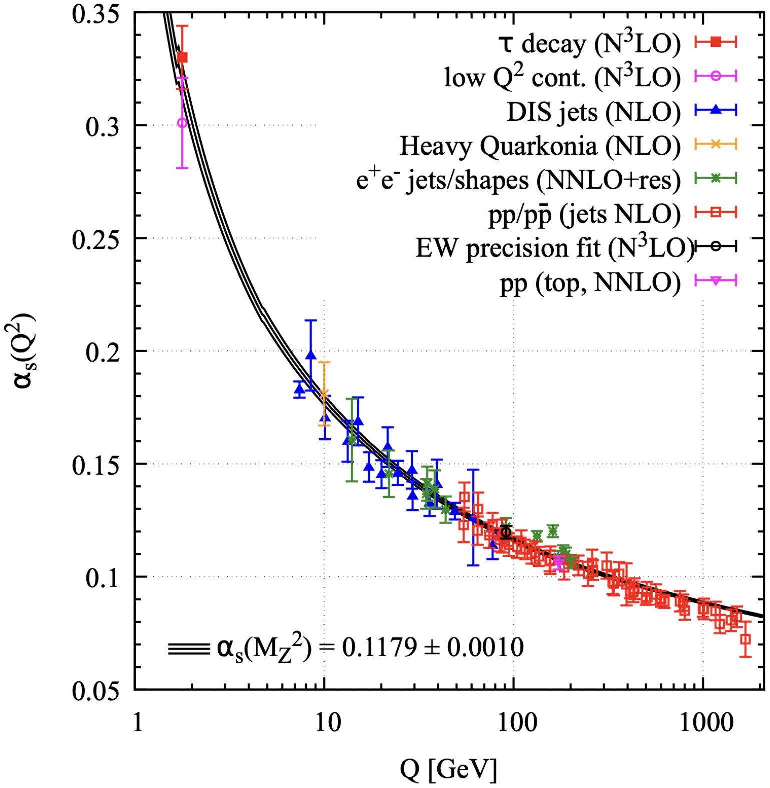


Figure 1.3: Summary of measurements of α_S as a function of the energy scale Q from various type of interactions and with different degrees of QCD perturbation theory in the extraction of α_S (NLO : next-to-leading order; NNLO : next-to-NLO; NNLO+res : NNLO matched to a resummed calculation; N^{3}LO : next-to-NNLO) [9].

A parton emitted at high energy is first in a state corresponding to the asymptotic freedom described previously, i.e it is able to propagate freely. During its propagation, it will loose energy by the emission of gluons or by the creation of quark - anti-quark pairs, that then can emit further partons : this process is named parton shower and is similar to the emission of photons by accelerated charged particles through the electromagnetic interaction. Once the energy of the partons is below the QCD energy scale, these partons will then gather into colorless hadrons during a step called the hadronization. It results in the formation of a narrow cone of hadronic particles that propagate in a direction close to the initial parton : this cone is called a hadronic jet.

In addition to the hard scattering process, interactions between other partons that come from the same hadrons and that involve less energy can happen in the same collision. These interactions are named «underlying event». Moreover, the partons belonging to the colliding hadrons can radiate partons before the interaction occurs : this process is called «initial state radiation» (ISR). After the interaction, partons in the final state can also emit other partons : this is called final state radiation (FSR). An illustration of a hadron-hadron collision with a hard scattering process and underlying event is shown in Figure 1.4 [10].

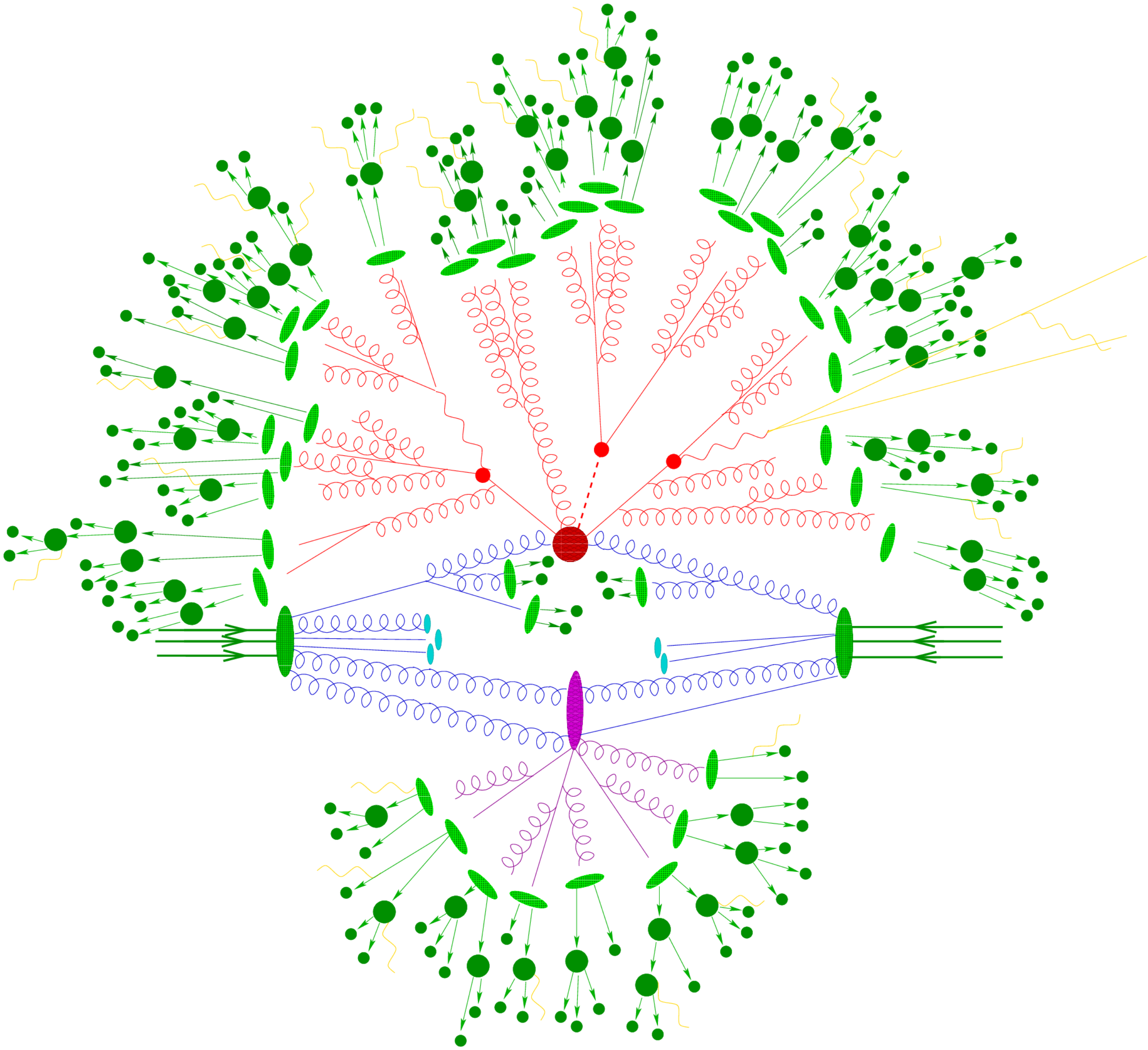


Figure 1.4: Illustration of a hadron-hadron collision. The red circle represents the hard scattering process of the collision, and the red tree-like structure is a representation of the parton shower. The green spots represent the hadronization process and the decay of the hadrons. The yellow lines represent the emission of photons by charged hadrons. The purple spot corresponds to the underlying event [10].

Chapter 2

The Dark Matter

Contents

2.1 Observational evidence

The dark matter is an invisible and hypothetical form of matter impacting visible matter through gravitational effects and possibly through other weakly-coupled interactions, making difficult its detection in laboratory. According to cosmological observations, it would compose roughly 85% of the total matter in the universe.

The first evidence of the presence of an invisible matter in the Universe was obtained by Fritz Zwicky in 1933 [11]. He considered the Coma Cluster of galaxy and estimated its mass based on the velocity dispersion of the galaxies near its extremity, and compared it to an estimation based on its luminosity and number of galaxies. He found that to explain the velocities of the galaxies, the cluster must contain about a few hundred times more mass than what is visible, otherwise the gravity effect of the visible galaxies would not be sufficient to maintain the galaxies bound together given their large velocities. He proposed the existence of an invisible matter in the cluster that would allow to explain this discrepancy.

Then, one of the most important evidence for dark matter came from the measurement of galaxy rotation curves by Vera Rubin and Kent Ford in the 1970s [12]. These curves present the orbital speeds of visible stars or gas in a given galaxy in function of the radial distance from the center of the galaxy. The Newton laws of mechanics predict that at large distance from the galaxy center, this orbital speed $v(r)$ should vary as $v(r) \sim 1/\sqrt{r}$, where r is the radial distance of the star or gas to the galaxy center. However, the rotation curves measured by Rubin and Ford present an unexpected behavior regarding the prediction from the Newtonian mechanics : the orbital speeds reach a constant value at large radii in the galaxy. These results have been confirmed in the following decades, and they are illustrated in Figure 2.1 with recent data [13].

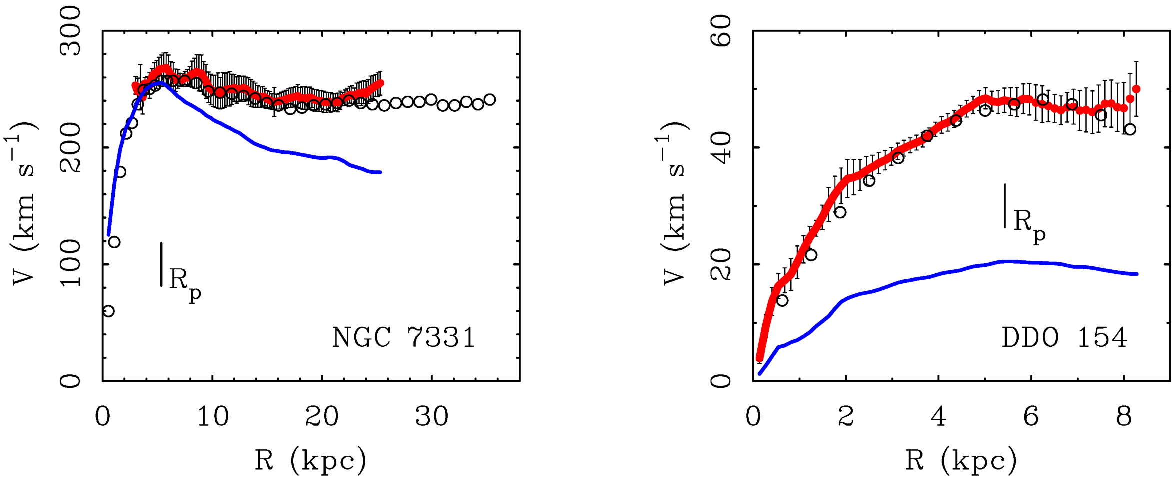


Figure 2.1: Rotation curves of two galaxies, NGC 7331 and DDO 154, observed with data from the THINGS survey (as red points with error bars, and represented as open points for older data). The blue lines show the rotation velocity coinciding with the observed stars and gas, and R_p represents the radius at which this blue curve is maximum [13].

These observations imply that the visible matter gravitational presence is not sufficient to explain these curves. Thus, there may exist a non-luminous matter that has a very different repartition in galaxies than the visible matter. The main corresponding hypothesis is to suppose the existence of a dark matter halo that surrounds each galaxy.

Another manifestation of the presence of dark matter in the universe can be observed through the phenomena of gravitational lensing. It is a consequence of the Einstein's theory of general relativity and it happens when a massive astrophysical object is located between a source of light and an observer. Because of the curvature of spacetime that it induces, the presence of this mass will act as a lens and curve the path of the emitted light from the source, so the observations will be distorted images of the distant object. For example, a galaxy cluster will act as a gravitational lens if it is placed between a source like a quasar and an observer. By measuring the distortion that is observed, the mass of the galaxy cluster can be deduced. In this way, the measurements of such effects can be used to probe the matter (both visible and invisible) distribution in galaxy clusters and to constrain the dark matter density.

The Cosmic Microwave Background (CMB) also provides an important evidence for the presence of dark matter, precisely measuring its density in the universe. Indeed, the CMB is very similar to a black-body radiation corresponding to a temperature of $T = 2.725$ K but it is not perfectly isotropic and presents extremely small fluctuations of temperature, related to the interactions between photons and matter in the very early universe. These anisotropies can be decomposed into an angular power spectrum that contains several peaks of different heights as measured by the Planck telescope (cf Figure 2.2 [14]).

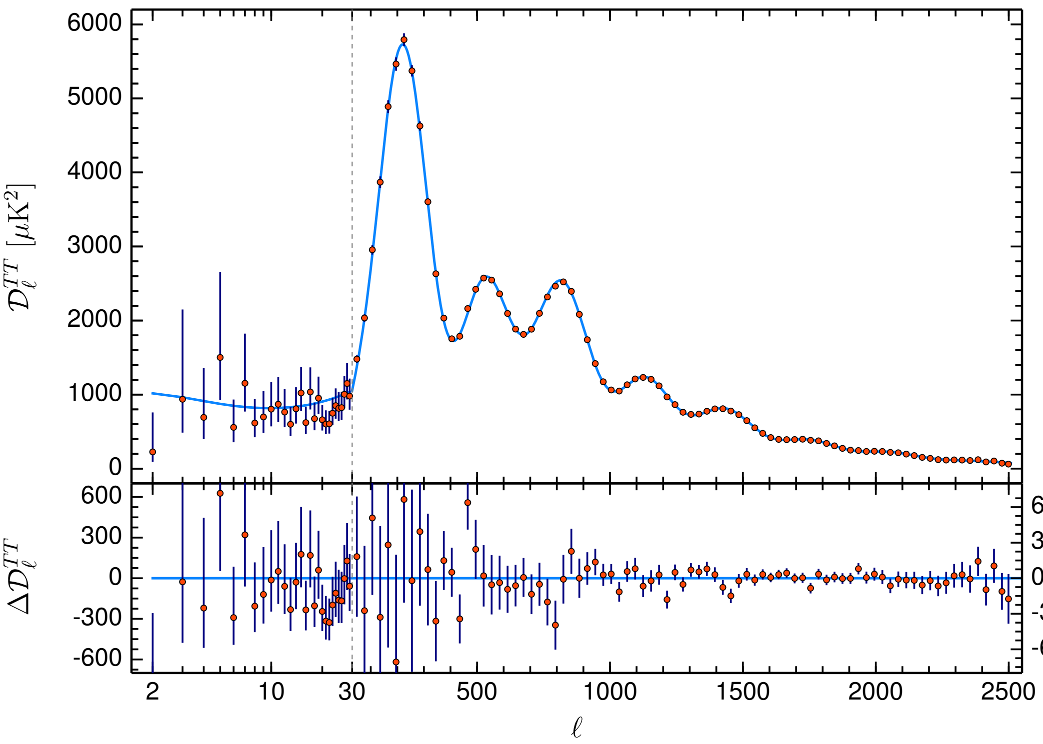


Figure 2.2: Angular power spectrum of the CMB temperature as measured by Planck in 2018. The red points correspond to data and the blue line is the result of a fit assuming the Lambda Cold Dark Matter model [14].

These peaks are related to cosmological parameters, such as the curvature of the universe, the density of baryonic matter and dark matter. In Figure 2.2 [14], the Lambda Cold Dark Matter model is used to fit this spectrum and to extract constraints on these cosmological parameters, providing the measured values of the ordinary and dark matter densities in the universe.

2.2 Dark matter hypotheses

In order to explain these observations, different hypotheses have been proposed.

The first conjecture was that dark matter could be massive astrophysical objects that would still be made up of baryonic matter but not emitting light (or not sufficiently to be detected). These objects are named MaCHOs for Massive Compact Halo Objects, and could correspond for instance to black holes or neutrons stars that may constitute the expected halo of Dark Matter.

However, several observations have demonstrated that the MaCHOs are not sufficient to explain the nature of dark matter, and can only contribute to a small fraction of the missing mass. For example, the presence of a MaCHO can be detected while passing between a star and an observer, bending the light that is emitted by the star and causing a detectable change in its brightness. This phenomenon is named gravitational microlensing. Searches for this kind of events have allowed to exclude a very large range of possible masses for MaCHOs.

Moreover, the CMB measurements have demonstrated that Dark Matter must be composed of non-baryonic matter. One of the most prevalent hypothesis is that dark matter could be made of new massive elementary particles that would interact with ordinary matter through gravity and possibly through other interactions that are as weak or weaker than the weak interaction. These sorts of particles are designated as Weakly Interacting Massive Particles (WIMPs).

They must satisfy some properties in order to correspond to the astrophysical and cosmological observations : the WIMPs must be massive and cold (i.e having large mass compared to Standard Model particles, and having non relativistic velocity, forcing the WIMPs to group themselves due to the gravitational interaction and to form large massive structure) and stable relatively to the age of the universe, because dark matter was present in the early universe and still is nowadays. Moreover, in order to explain the current dark matter density, the self-annihilation cross-section between dark matter particles and anti-particles cannot be higher than the weak interaction processes cross-section.

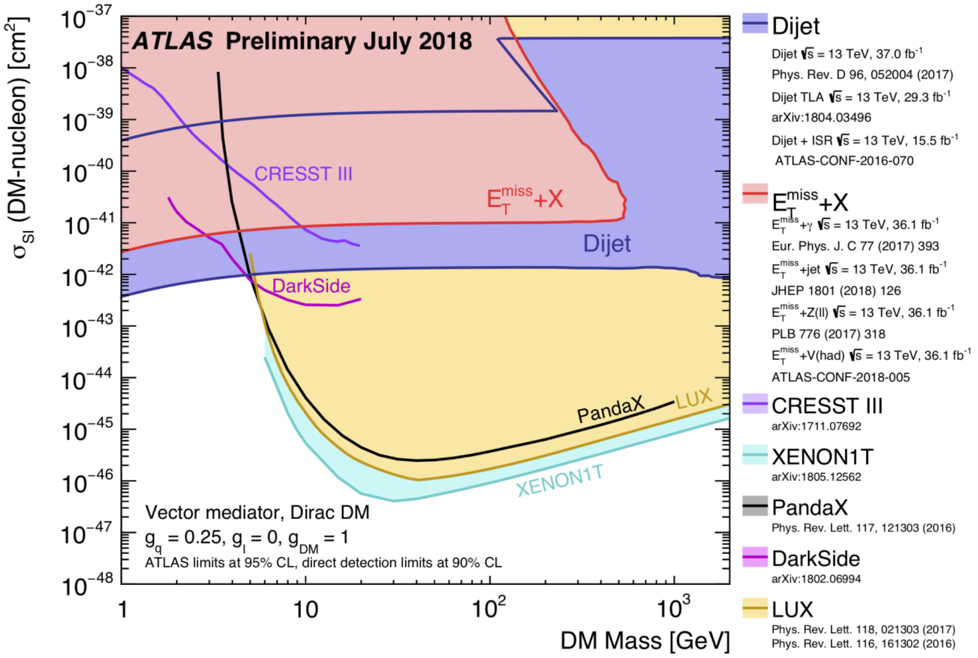


Figure 2.3: Comparison between constraints on the WIMP-nucleon scattering cross-section as a function of the WIMP mass obtained from direct detection experiments and searches at the LHC (more details on the different methods to detect Dark Matter are presented in section 2.3). The shaded areas correspond to the phase-space regions that are excluded [15].

Many Beyond-the-Standard Model (BSM) theories predict the existence of WIMPs, with supersymmetry being the most popular. However, none of these theories have been confirmed by experiments that were specifically searching for WIMPs. Nowadays, strong constraints on the WIMPs mass and on the WIMPs-baryonic matter scattering cross-section have been set as shown in Figure 2.3 [15], reducing significantly the allowed regions of the model parameter space.

The WIMPs hypothesis is not totally rejected but in addition to it, several other dark matter theoretical models involving new particles have been proposed and tested by experiments, such as the axions or the existence of a dark sector. The latter will be more detailed in section 2.4 since it is related to the studies presented in this thesis.

2.3 Detection methods

If dark matter is made of new particles, three different and complementary experimental ways to detect them can exist, as illustrated in Figure 2.4 [16].

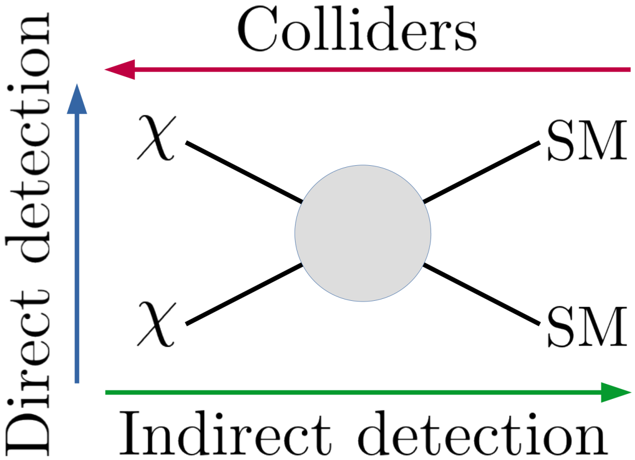


Figure 2.4: Illustration of the different methods to detect Dark Matter particles [16].

First, the direct detection aims to observe recoils of nuclei in laboratory due to interactions with particles of dark matter. Indeed, it is expected that a dark matter halo surrounds the Earth and so this latter should be permanently crossed by particles of dark matter. The recoil of the interacting nucleus will cause the emission of a small amount of energy that then can be detected. In order to increase the sensitivity of particle detection, cryogenic detectors can be employed, aiming to detect the heat produced by the recoil of a nucleus in a crystal absorber. Noble liquid detectors can also be used, relying on the detection of scintillation light and ionization signal that would be produced when a dark matter particle interacts in a volume of liquid such as xenon or argon.

One of the main concerns of this method is to highly reduce the potential background coming from Standard Model particles interactions (which can be done by locating such experiments in underground laboratories to lower the contribution from cosmic particles and using radio-pure materials) but also to discriminate the remaining background collisions from the ones induced by dark matter particles.

Secondly, the indirect detection is searching for Standard Model particles that may come from the self-annihilation of dark matter particles happening in high-dark-matter-density regions. These annihilation products may be gamma rays or pairs of Standard Model particle - anti-particle. In general, the indirect detection consists in observing positrons, anti-protons or gamma rays coming from high-density regions with telescopes and searching for excesses, which would be a potential manifestation of dark matter self-annihilation.

Finally, the last approach to detect dark matter particles is to produce them through very rare processes by colliding Standard Model particles at high energy, such as in hadron-hadron collisions at the LHC. However, if such particles are produced, they will cross the detectors without interacting because dark matter particles would interact weakly with ordinary matter. They can only be observed indirectly if they are produced together with Standard Model particles. Indeed, because of the law of conservation of momentum, the sum of the momentum transverse to the collisions of all the particles in the final state must be zero. For an event where invisible particles have been produced with visible ones, one can only detect the contribution from the latter and the contribution from the invisible particles will appear as a large missing transverse momentum needed to ensure the momentum conservation.

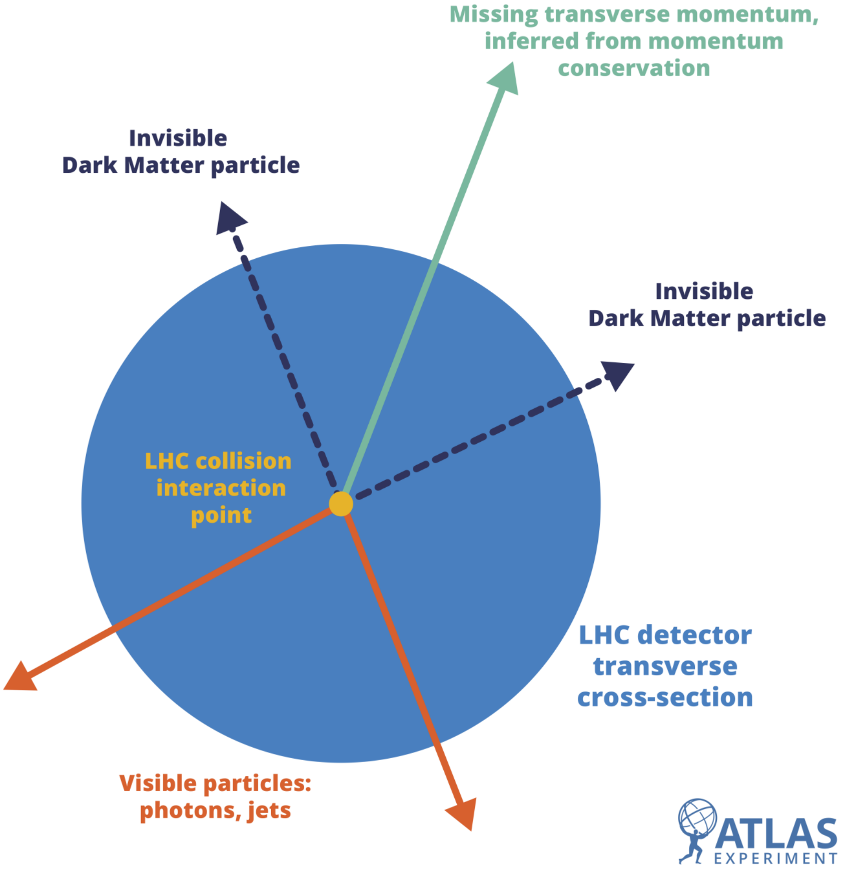


Figure 2.5: Scheme illustrating the production of dark matter particles in association to visible ones at the LHC [15].

This type of signature corresponding to a large missing transverse momentum balancing visible objects (that can be hadronic jets, leptons or photons that could come either from the process of dark matter particles production or from ISR) is illustrated in Figure 2.5 [15]. The Feynman diagram of a BSM process that may reproduce this signature is shown in Figure 2.6 [17].

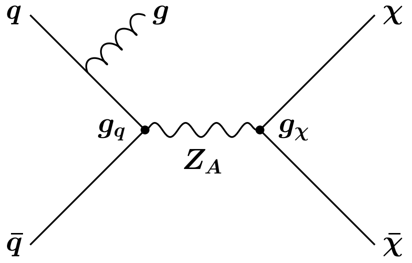


Figure 2.6: Diagram of the pair production of WIMPs χ through a new mediator Z_A with initial state radiation [17].

The current searches for dark matter particles at colliders are also looking for other types of final states, as will be described in the next section.

2.4 Dark sector

Dark matter theoretical models can also be based on the «Asymmetric Dark Matter» (ADM) paradigm which is totally different from the WIMPs one. ADM models are motivated by the observation that the mass density of dark matter, Ω_{DM}, is similar to the one of baryonic matter Ω_{b} ($\Omega_{\mathrm{DM}} \simeq 5\Omega_{\mathrm{b}}$). The current mass density of baryons is the result of the matter - anti-matter annihilation process in the early universe; the slight excess of baryons is the matter seen in the Universe today. The ADM paradigm postulates that the current mass density of dark matter should also be due to a dark matter - anti-dark-matter annihilation happening in the early universe that may also have presented an asymmetry. In this way, the baryonic and dark matter asymmetries could share a common origin and be the results of production processes that have been connected in the early universe.

As the components of baryonic matter correspond to stable states of the Standard Model sector, the ADM paradigm suggests that Dark Matter could also be constituted of stable elements of a dark sector that may be described by a new gauge group containing new quantum fields. In particular, this new gauge group may corresponds to a $SU(N_{C_d})$ gauge group describing a QCD-like theory characterizing this dark sector : the latter is generally named dark QCD. This dark QCD theory describes the interaction between new quantum fields referred as dark quarks q_d and dark gluons g_d in a similar way as the QCD describes the interactions between Standard Model quarks and gluons. It is characterized by free parameters like the number of dark colors N_{C_d}, the number of flavors of dark quarks N_{f_d}, a coupling constant α_d or by a confinement scale Λ_d which plays the same role as the QCD confinement scale. Thus, dark quarks emitted at high energy can produce dark jets containing composite dark hadrons during a process similar to the one that Standard Model quarks realize.

In such a theory, it is also supposed that dark QCD particles can be connected to Standard Model ones through a new interaction which can be mediated by a new vector boson Z' or a new scalar mediator ϕ for instance. Thus, at hadron colliders like the LHC, it would be possible to directly produce dark quarks through this new interaction and initiate the formation of jets containing dark hadrons. The latter can either be stable within the dark sector (and so remain invisible) or unstable and decay back to Standard Model particles through the same mediator. Depending on the fraction of invisible dark hadrons inside the dark jets and on the decay life-time of the unstable ones, the Standard Model particles coming from the decay of the unstable dark hadrons will produce very unusual and exotic signatures in the detectors that are generally classified into three categories as schematized in Figure 2.7 [18] :

- Dark jets that correspond to the case where the dark hadrons are all unstable and promptly decay back to Standard Model particles (QCD-like region in Figure 2.7).
- Semi-visible jets (SVJ) that consider the case where a fraction of dark hadrons do not decay to Standard Model particles and remain invisible, leading to a large missing transverse momentum in the final state which is aligned with one of the jets (Exotic (II) in Figure 2.7).
- Emerging jets (EJ) that result from the case where the dark hadrons are all unstable but decay to Standard Model particles with an important life-time, resulting in detectable signals that are displaced from the interaction point (Exotic (I) in Figure 2.7).

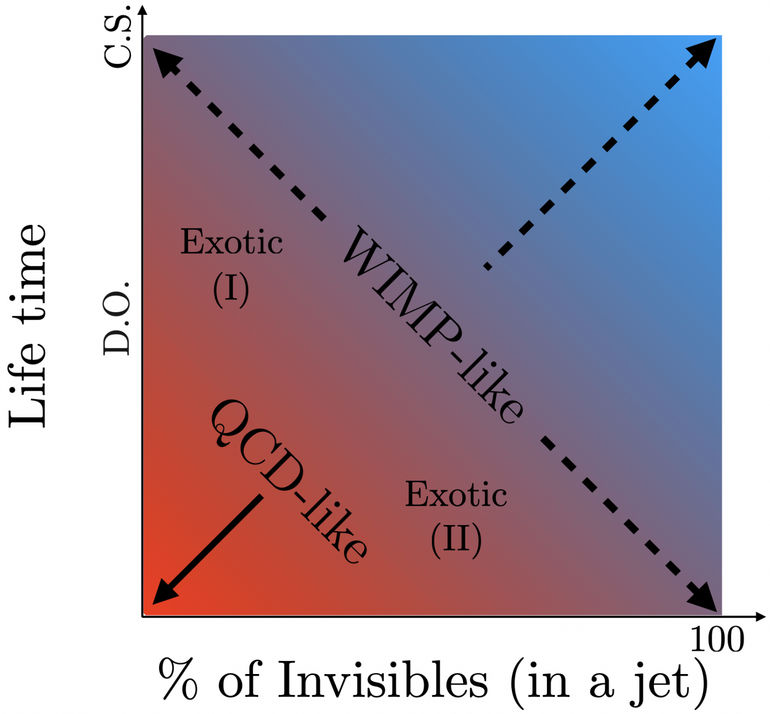


Figure 2.7: Diagram representing the different topologies of jets from dark QCD. D.O. means «Displaced Objects» and C.S. corresponds to «Collider Stable» [18].

These type of jets also differ from QCD ones in terms of internal jet structure. Indeed, if the produced visible particles are Standard Model quarks, the latter will undergo parton shower and hadronization in the visible sector. In this way, the jets from the dark sector can be the result of a double hadronization process : one with the dark quarks in the dark sector, followed by one in the visible sector.

In Figure 2.7, there is also a «WIMP-like» region where either the unstable dark hadrons lifetime is too large regarding the detector size, or the fraction of invisible particles is close to one. The resulting final state is equivalent to the one coming from a pair production of WIMPs described in section 2.3, so this area is already covered by WIMP searches at colliders. Thus, searches for dark, semi-visible and emerging jets at colliders are complementary to the WIMP searches that have been previously performed.

As it is the main subject of the studies presented in this document, the emerging jet model will be detailed in section 5.1.

Chapter 3

The LHC and the ATLAS experiment

Contents

3.1 The LHC

The LHC is a particle collider located at CERN near Geneva across the border of France and Switzerland. It represents the most powerful particle collider regarding the energy of collision and consists in a circular accelerator with a circumference of 27 kilometers situated in a tunnel about 100 meters underground. It is able to produce both proton-proton (*pp*) and heavy-ion collisions, but only the former will be considered in this thesis.

Before being injected in the LHC, the protons are progressively accelerated to injection energy thanks to a chain of accelerators that is presented at Figure 3.1 [19]. First, negative hydrogen ions, i.e hydrogen atoms with an additional electron, are accelerated by the Linear Accelerator 4 (LINAC4) to the energy of 160 MeV. The protons composing the ions are then injected in the Proton Synchrotron Booster (PSB) and accelerated to 2 GeV before entering the Proton Synchrotron (PS), the latter allowing to increase the proton beam energy to 26 GeV. Finally, the protons are injected to the Super Proton Synchrotron (SPS) and accelerated to 450 GeV before entering the two beam pipes of the LHC : one beam is circulating clockwise while the other one is circulating anti-clockwise.

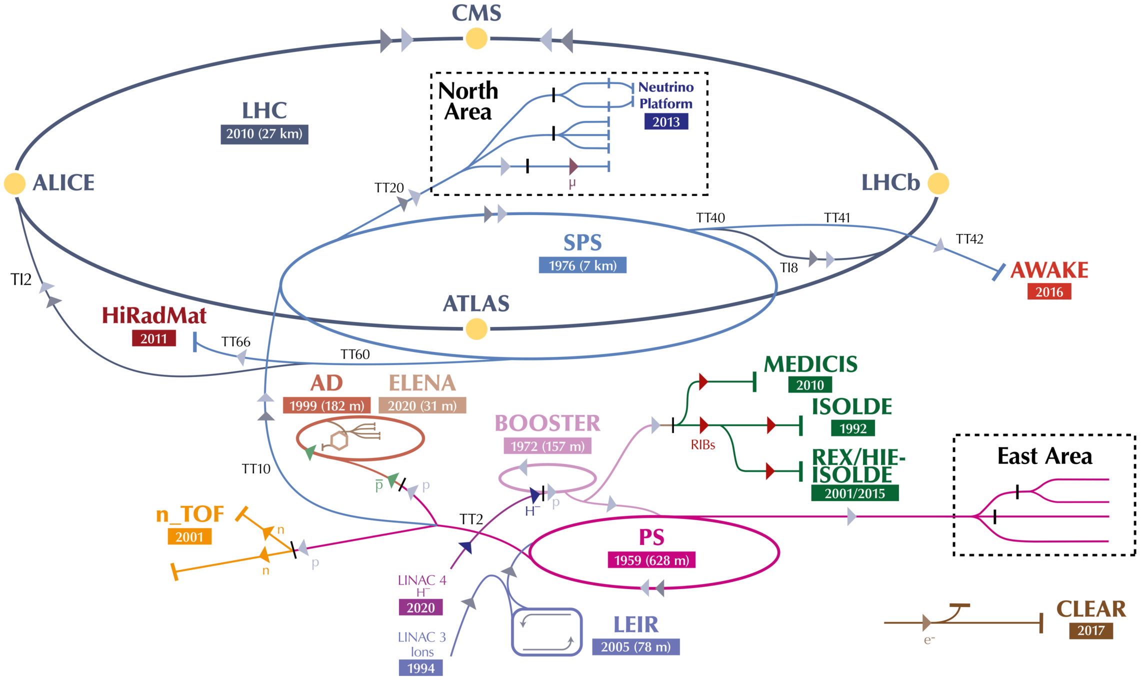


Figure 3.1: The CERN's accelerator complex [19].

Both beams are contained in tubes kept at ultrahigh vacuum and are guided around the accelerator ring by magnetic fields generated by superconducting magnets employed at extremely low temperature. In the LHC, the protons are accelerated with the use of radio-frequency cavities in order to reach their maximum energy : it is currently equal to 6.8 TeV and thus leads to collisions with a centre-of-mass energy equal to 13.6 TeV. During their circulation around the LHC, the protons are actually grouped together into separated bunches (about 2800 effective bunches in each beam), each one containing about 10^{11} protons.

There are four interaction points at the LHC where the two beams are collided, each point being surrounded by a particle detector. These four detectors are :

- ATLAS [20] and CMS [21] which are optimized for *pp* collisions and aim to study the Standard Model and explore BSM theories,
- ALICE [22] which is dedicated to the study of heavy-ions collisions,
- and LHCb [23] which is notably used to study the physics of the bottom quark.

During data taking periods, a collision at the interaction points between two bunches occurs every 25 ns, i.e at a rate of 40 MHz.

The first data taking period named Run-1 started in 2010 and ended in 2013 : during this period that has been marked by the discovery of the Higgs boson, the centre-of-mass energy has been increased from 7 to 8 TeV. The LHC has then been shut down for two years in order to increase the centre-of-mass energy notably. Then, the Run-2 data taking period took place between 2015 and 2018 and involved a collision energy of 13 TeV. After a second shutdown, the Run-3 started in 2022, recording data corresponding to a centre-of-mass energy of 13.6 TeV. This thesis will present results that are using *pp* collisions data collected in 2022 and 2023.

3.2 The ATLAS detector

ATLAS, which is the acronym for «A Toroidal LHC Apparatus», is a general purpose particle detector located at the LHC. As CMS, it has been designed both to accurately measure the Standard Model properties and to be able to eventually confirm BSM theories. A global view of the detector is presented in Figure 3.2 [24].

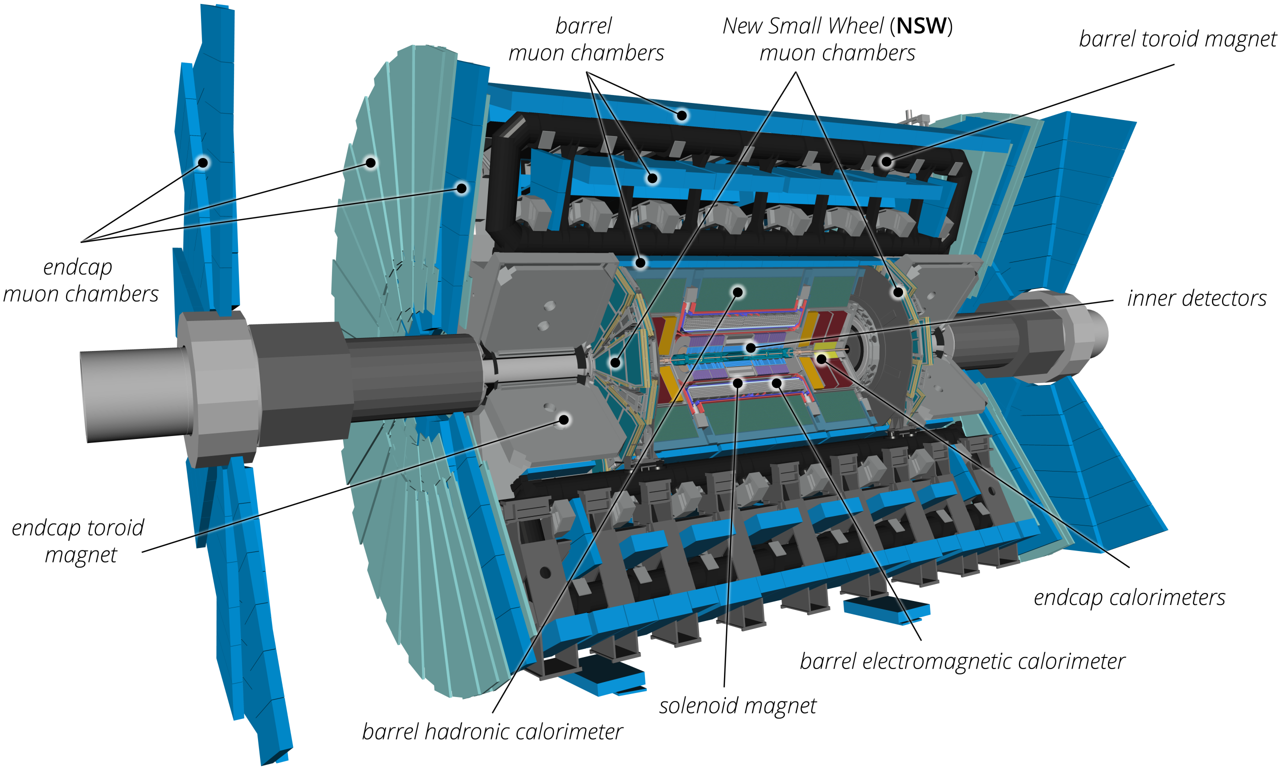


Figure 3.2: The ATLAS detector [24].

ATLAS has a cylindrical shape of dimensions 44 meters in length and 25 meters in diameter, and weights approximatively 7000 tonnes. The detector enables to cover nearly the entire solid angle around its center. It is designed in layers that make complementary measurements, providing accurate measurements of the particles trajectory, energy and momentum and allowing their identification.

The main components of the detector, that will be described in the following sections, are the inner detector, the calorimeters, the muon spectrometer and the magnet system.

3.2.1 Coordinate system

The coordinate system used by ATLAS is illustrated in Figure 3.3 [25].

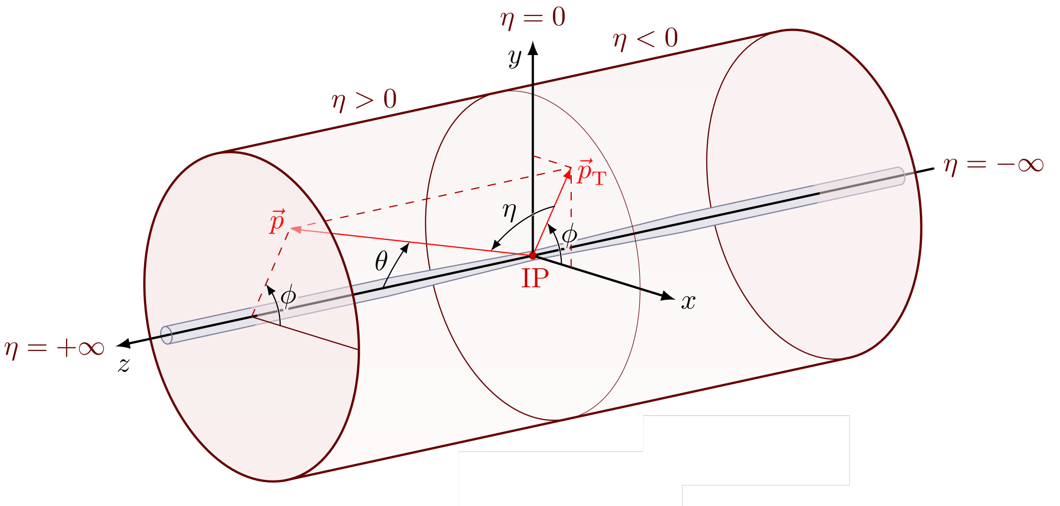


Figure 3.3: Coordinate system used in ATLAS [25].

The x-axis is pointing towards the center of the LHC ring, the y-axis is directed upwards and the z-axis is aligned with the beam, oriented in the direction of the one that is circulating anti-clockwise. The origin of the coordinate system corresponds to the center of the detector.

A position can be defined by the three parameters (r, θ, ϕ), with r the distance to the origin, θ the polar angle measured with respect to the beam and ϕ the azimuthal angle around the z-axis. The momentum vectors can be decomposed into a contribution in the transverse (x, y) plane, p_T, and a contribution along the direction of the beam p_Z.

Rather than θ, it is the pseudo-rapidity, defined as $\eta = -\ln(\tan(\theta/2))$, that is generally used as coordinate. In this way, $\theta = \pi/2$ is equivalent to $\eta = 0$, while $\theta = 0(\pi)$ corresponds to $\eta = +\infty(-\infty)$. The pseudo-rapidity can be written as a function of the particle three-momentum $\vec{p}$ as $\eta = (1/2)\ln((|\vec{p}|+p_z)/(|\vec{p}|-p_z))$. To evaluate the distance between two points, it is the angular distance in the plane (η, ϕ) defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ that is generally used.

Instead of η, the rapidity y, defined as $y = (1/2)\ln((E+p_z)/(E-p_z))$ can also be used. Indeed, compared to the pseudo-rapidity, the advantage is that the difference between the rapidities of two particles is invariant with respect to Lorentz boosts along the beam axis. Moreover, in the case of a highly relativistic particle, the rapidity is equivalent to the pseudo-rapidity because $E \approx |\vec{p}|$.

3.2.2 Magnet system

The magnet system aims to produce an important magnetic field in both the inner detector and the muon spectrometer to curve the trajectory of the charged particles produced in collisions. Indeed, once the particle trajectory is reconstructed, the direction and the radius of curvature can respectively provide the charge and the momentum of the particle. The magnet system is divided in two parts that are shown in Figure 3.4 [26] : the central solenoid magnet and the toroid ones.

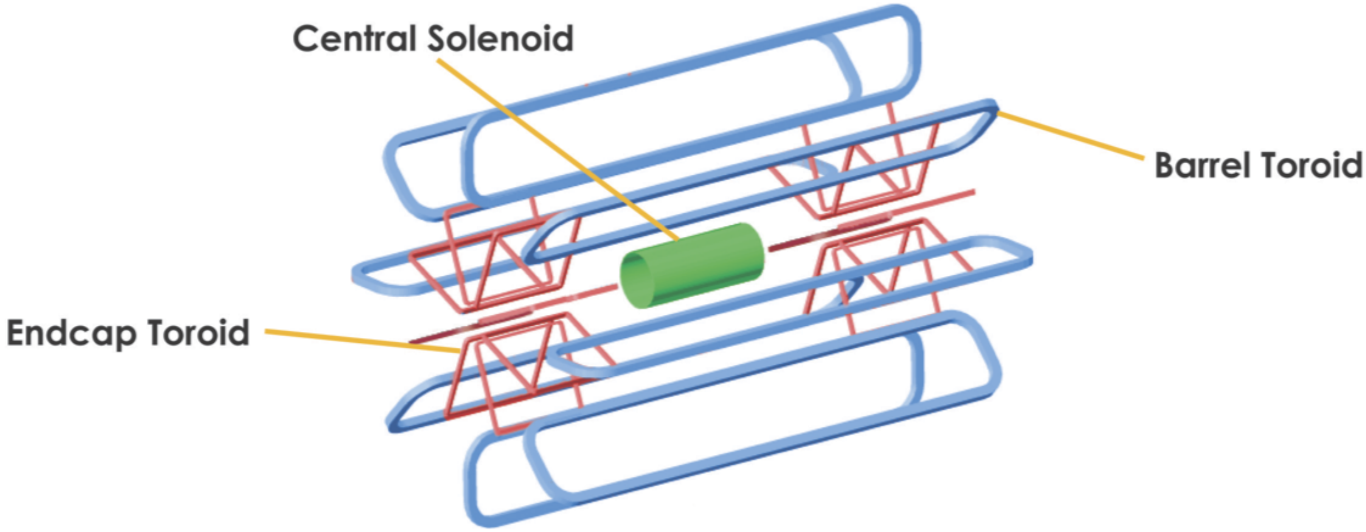


Figure 3.4: Magnet system in ATLAS [26].

The central solenoid magnet, which surrounds the inner detector, is 5.8 meters long, has a diameter of 2.6 meters and weighs 5 tonnes. Its thickness is only 4.5 cm. It is made of a superconducting Niobium-Titanium alloy employed at extremely low temperature ($\sim$ 5K), and it is able to produce a 2T axial magnetic field in the inner detector.

The toroid magnet system consists of three magnets also made of Niobium-Titanium : one in the center of the detector (barrel toroid) and one in each extremity (endcap toroids). The barrel toroid magnet is 25.3 meters long and has a 20.1 outer diameter. The end-cap toroids are 5 meters long and have a 10.7 meters outer diameter. Each toroid uses eight separate air-core coils to produce the magnetic field in the muon spectrometer.

The solenoidal and toroidal magnetic field lines are illustrated in Figure 3.5 [27].

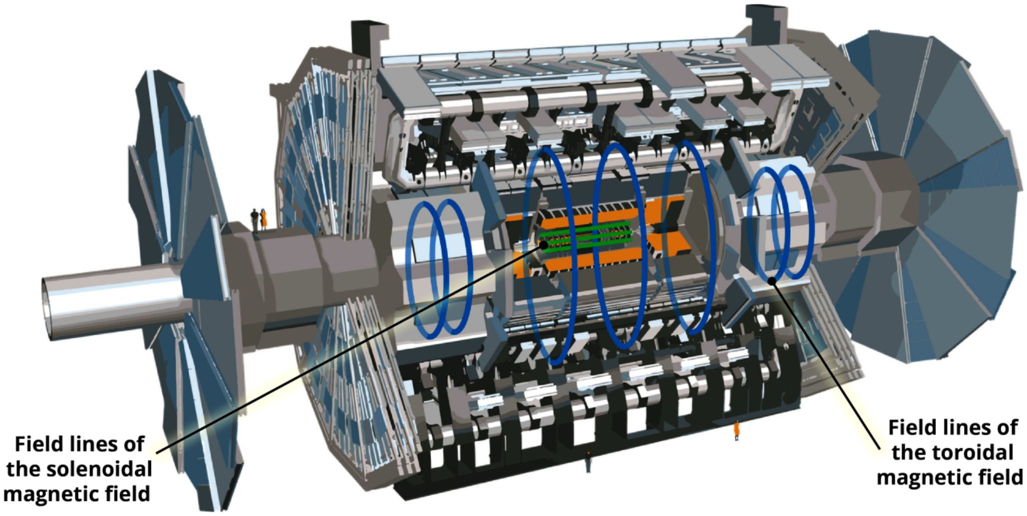


Figure 3.5: Solenoidal (in green) and toroidal (in blue) magnetic field lines [27].

3.2.3 Inner detector

The innermost ATLAS sub-detector is named inner detector : it is the first component of the detector crossed by emitted particles and it is shown in Figure 3.6 [24]. As mentioned previously, it is immersed in a 2T axial magnetic field that allows to curve the trajectory of the charged particles. A charged particle crossing the inner detector will deposit energy in the different elements : these energy deposits are named hits and are used together to reconstruct a track, corresponding to the charged particle trajectory (the track reconstruction will be detailed in section 4.1).

The inner detector is divided into three sub-components, which are the Pixel detector, the Semi-Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT). Their disposition in the transverse plane is schematized in Figure 3.7 [24]. They allow to cover the pseudo-rapidity range $|\eta| < 2.5$ while having a full coverage in ϕ.

Pixel detector The pixel detector is the first component of the inner detector that the produced particles will cross. It is composed of four concentric cylindrical layers of semiconductor silicon pixels in the barrel region located at radii of 33.25, 50.5, 88.5 and 122.5 mm, and of three disks of silicon pixels perpendicular to the beam line that are located in each of the two end-cap regions at $|z|$ of 495, 580 and 650 mm.

The layer pixel sensors are segmented transversely and longitudinally, i.e in $r\phi$ and z respectively : the innermost layer pixels have a transverse size of 50 μm and a longitudinal one of 250 μm while the three outermost layers pixels have a transverse size of 50 μm and a longitudinal one of 400 μm. The accuracies in the barrel are 10 μm in $r\phi$ and 115 μm in z, while in the end-cap, they are 10 μm in $r\phi$ and 115 μm in r. In the pixel detector, the number of readout pixel channels is 92 million.

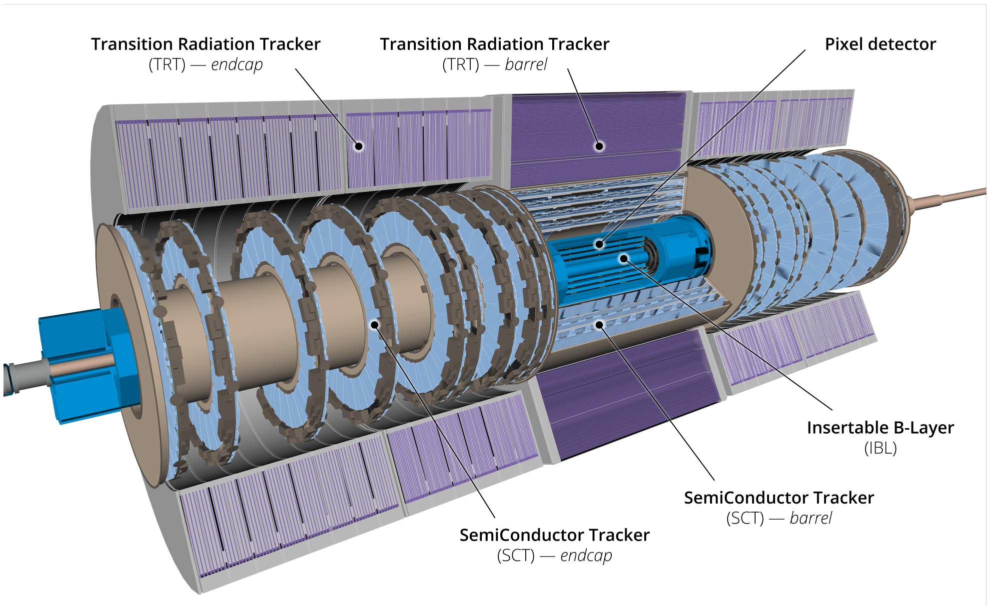


Figure 3.6: Overview of the inner detector [24].

When a charged particle crosses a pixel sensor, it generates free charge carriers that are then collected at electrodes through the application of an electric field. It results in a current pulse that can be read-out by associated electronics and recorded as a hit, providing a three-dimensional point located along the particle trajectory. Generally, for a single charged particle, four hits in the pixel detector will be recorded.

SCT The SCT directly surrounds the pixel detector. It is composed of four concentric cylindrical layers of silicon strips in the barrel region located at radii of 299, 371, 443 and 514 mm, and of nine perpendicular disks of silicon strips in each of the two end-cap regions.

In the barrel, each layer uses 40 mrad stereo angle strips in order to measure $r\phi$ and z, and strips parallel to the beam direction to measure $r\phi$ while in the end-cap regions, each disk uses strips running radially and strips at an angle of 40 mrad. The mean pitch of a strip is approximatively 80 μm. The accuracies in the barrel are 17 μm in $r\phi$ and 580 μm in z, while the ones in the disks are 17 μm in $r\phi$ and 580 μm in r. In the SCT, the total number of readout channels is about 6.3 million.

The physical process describing the interaction between a charged particle and the sensor is similar to the one described for the pixel detector. The SCT can typically provide four hits for a single charged particle.

TRT The outermost component of the pixel detector is the TRT. Unlike the precedent components, the TRT is made of 4 mm diameter drift tubes filled with a gas (a mixture of Xe, CO_2 and O_2) and having a 30 μm diameter gold-plated tungsten wire in its centre. When a charged particle crosses a tube, the gas is ionized and the produced electrons drift towards the central wire. The resulting signal is then read-out and recorded as an hit.

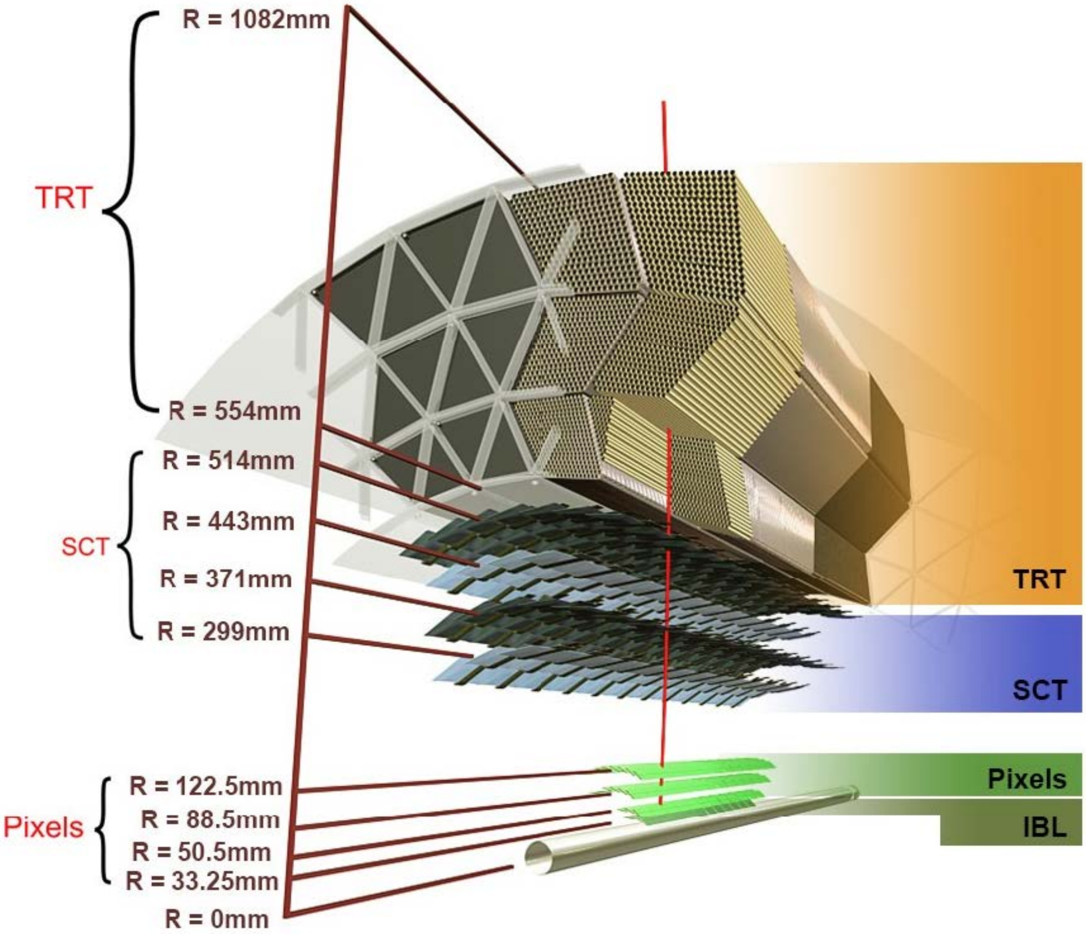


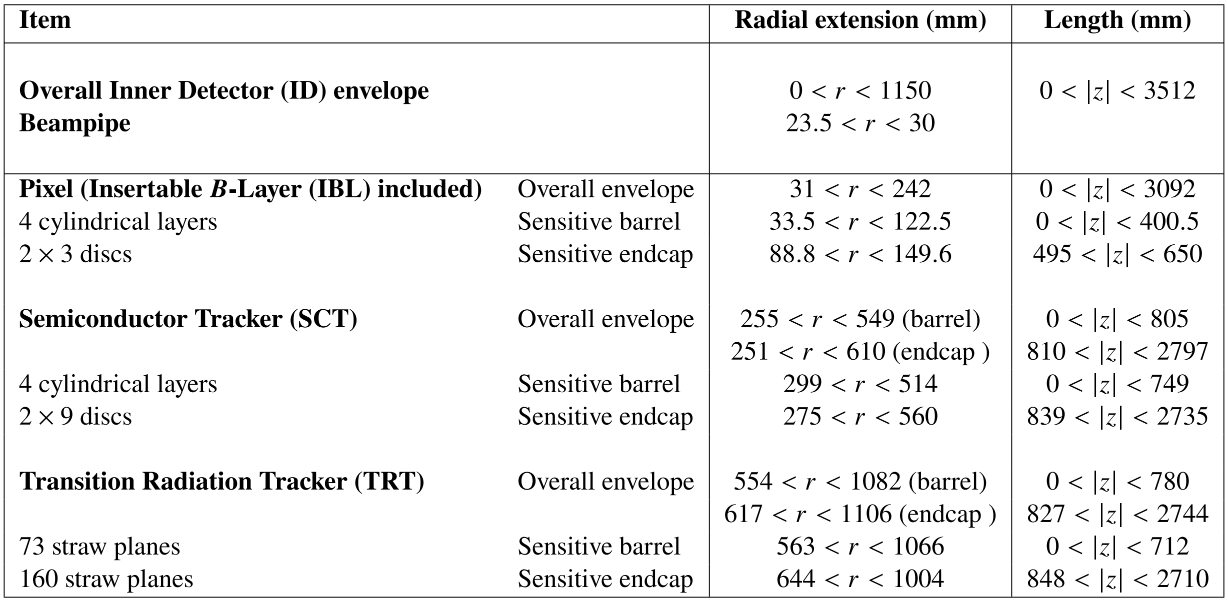
Figure 3.7: Transverse view of the inner detector [24].

The barrel contains approximatively 50 000 tubes that have a 144 cm length and that are arranged parallel to the beam at radii from 554 to 1082 mm. They provide measurement of r and ϕ with a resolution of 130 μm. On the other hand, each end-cap region contains approximatively 125 000 tubes that are perpendicular to the beam and that have a 39 cm length. They provide measurement of z and ϕ also with a resolution of 130 μm. The total number of readout channels in the TRT is approximatively 350 000.

The spaces between the tubes are filled with polymer fibres in the barrel and foils in the end-cap regions in order to create transition radiation that may be emitted by highly relativistic traversing charged particles. The emitted photons can be absorbed by Xe atoms, depositing additional energy in the gas and increasing significantly the readout signals.

Unlike the pixel detector and the SCT, the hits in the TRT correspond to two-dimensional points. In general, the TRT provides 36 hits for one charged particle. In this way, the poorer resolution for the TRT compared to the pixel detector and the SCT ones is compensated by the larger number of hits that it produces.

The radial extension and the length of each sub-components of the inner detector is presented in Figure 3.8 [24].



Item		**Radial extension (mm)**	**Length (mm)**
Overall Inner Detector (ID) envelope		$0 < r < 1150$	$0 < \|z\| < 3512$
Beampipe		$23.5 < r < 30$	
Pixel (Insertable *B*-Layer (IBL) included)	Overall envelope	$31 < r < 242$	$0 < \|z\| < 3092$
4 cylindrical layers	Sensitive barrel	$33.5 < r < 122.5$	$0 < \|z\| < 400.5$
2×3 discs	Sensitive endcap	$88.8 < r < 149.6$	$495 < \|z\| < 650$
Semiconductor Tracker (SCT)	Overall envelope	$255 < r < 549$ (barrel)	$0 < \|z\| < 805$
		$251 < r < 610$ (endcap)	$810 < \|z\| < 2797$
4 cylindrical layers	Sensitive barrel	$299 < r < 514$	$0 < \|z\| < 749$
2×9 discs	Sensitive endcap	$275 < r < 560$	$839 < \|z\| < 2735$
Transition Radiation Tracker (TRT)	Overall envelope	$554 < r < 1082$ (barrel)	$0 < \|z\| < 780$
		$617 < r < 1106$ (endcap)	$827 < \|z\| < 2744$
73 straw planes	Sensitive barrel	$563 < r < 1066$	$0 < \|z\| < 712$
160 straw planes	Sensitive endcap	$644 < r < 1004$	$848 < \|z\| < 2710$

Figure 3.8: Spatial extension of the inner detector components [24].

3.2.4 Calorimeter

The calorimeter directly surrounds the inner detector and the solenoid magnet. It aims to measure the energy of both charged and neutral particles produced in the collision (except muons and neutrinos) by stopping their progression and forcing them to deposit all of their energy.

The calorimeter disposition is presented in Figure 3.9 [24]. It is made of two main components : an inner part, the electromagnetic calorimeter, which is designed to stop photons and electrons, and an outer part, the hadronic calorimeter, which stops hadronic particles escaping the electromagnetic calorimeter. The electromagnetic calorimeter is divided into a barrel region (EMB) and two end-cap regions (EMEC). In the same way, the hadronic calorimeter is divided between a barrel region and end-cap components (HEC). Moreover, the calorimeter contains forward components (FCal) that allow to cover the range $|\eta| < 4.9$.

Both electromagnetic and hadronic calorimeters consist in an alternation of absorbing (or passive) layers that are made with high density material in order to absorb the incident particles and force them to initiate a particle shower, and of layers with an active material aiming to measure the energy contained in the shower. Because particles may deposit energy in the passive layers, the energy measured by the active layers do not correspond to the totality of the incoming energy.

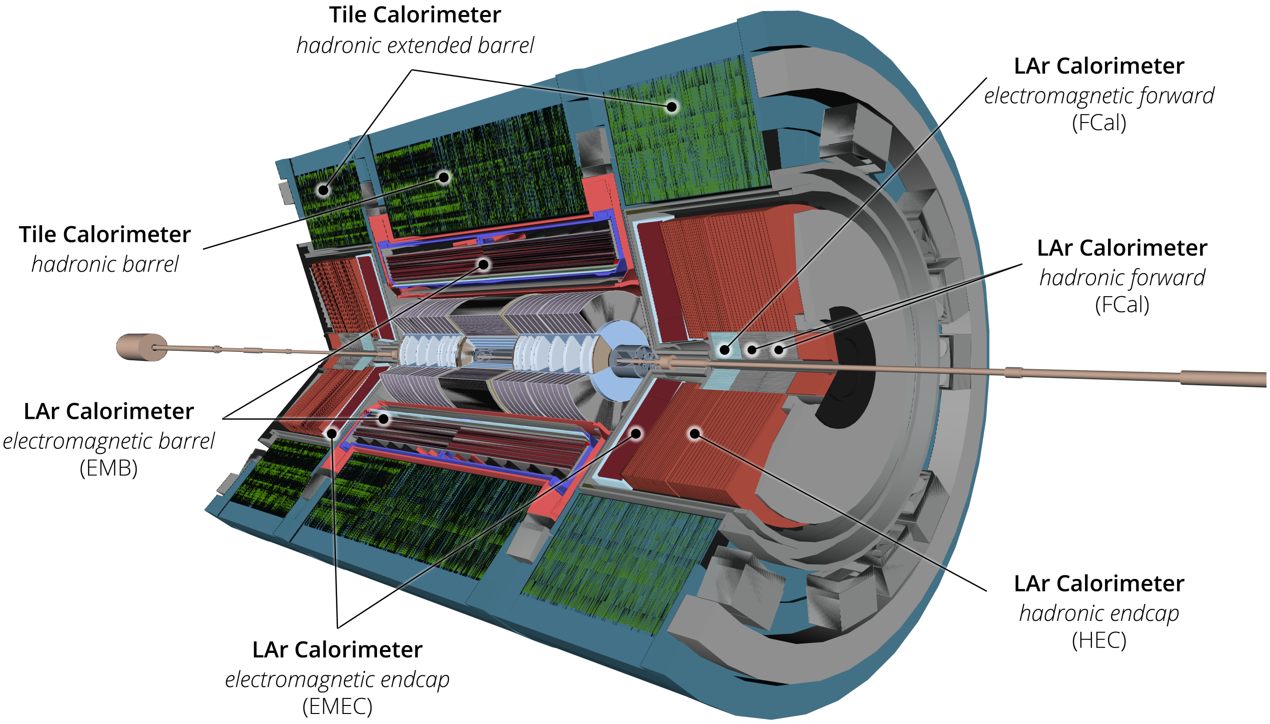
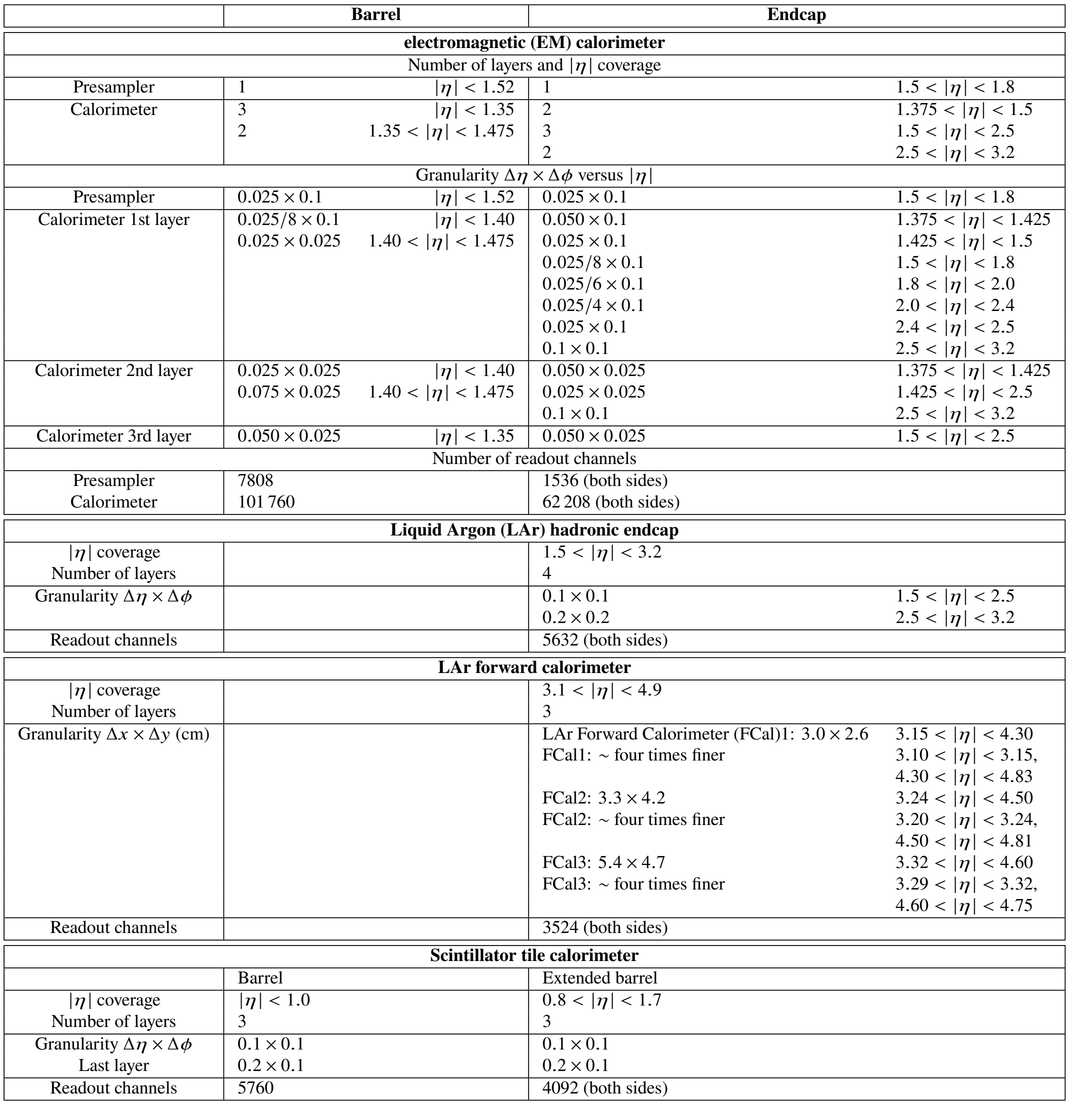


Figure 3.9: Overview of the calorimeter [24].

Two different technologies are used in the calorimeter concerning the material used in the active and passive layers. The first one is named «LAr Calorimeter» in Figure 3.9 [24], and corresponds to absorbing layers made of metal (either tungsten, copper or lead) and active layers made of liquid argon. The latter is ionized by incoming particles that produce a measurable current with an intensity proportional to the incoming particle energy. The second one is named «Tile Calorimeter» and corresponds to absorbing layers made of steel and active layers made of plastic scintillating tiles. As the particles forming the shower cross the plastic scintillators, the latter produce photons that are converted into a current that is then readout by the associated electronics.

The liquid argon is used in the electromagnetic calorimeter in the barrel (EMB) and end-cap regions (EMEC), in the hadronic calorimeter end-caps (HEC) and in the forward components (FCal). On the other hand, the tiles are used in the hadronic calorimeter in the barrel region. Each of these components is split into several cylindrical layers that are divided into cells with a certain granularity in the plane (η, ϕ) in order to be able to determine precisely the location of the energy deposits. The granularity and the η coverage of each layer is presented in Figure 3.10 [24].



	Barrel		Endcap	
electromagnetic (EM) calorimeter				
Number of layers and $\|\eta\|$ coverage				
Presampler	1	$\|\eta\| < 1.52$	1	$1.5 < \|\eta\| < 1.8$
Calorimeter	3	$\|\eta\| < 1.35$	2	$1.375 < \|\eta\| < 1.5$
	2	$1.35 < \|\eta\| < 1.475$	3	$1.5 < \|\eta\| < 2.5$
			2	$2.5 < \|\eta\| < 3.2$
Granularity $\Delta\eta \times \Delta\phi$ versus $\|\eta\|$				
Presampler	0.025×0.1	$\|\eta\| < 1.52$	0.025×0.1	$1.5 < \|\eta\| < 1.8$
Calorimeter 1st layer	$0.025/8 \times 0.1$	$\|\eta\| < 1.40$	0.050×0.1	$1.375 < \|\eta\| < 1.425$
	0.025×0.025	$1.40 < \|\eta\| < 1.475$	0.025×0.1	$1.425 < \|\eta\| < 1.5$
			$0.025/8 \times 0.1$	$1.5 < \|\eta\| < 1.8$
			$0.025/6 \times 0.1$	$1.8 < \|\eta\| < 2.0$
			$0.025/4 \times 0.1$	$2.0 < \|\eta\| < 2.4$
			0.025×0.1	$2.4 < \|\eta\| < 2.5$
			0.1×0.1	$2.5 < \|\eta\| < 3.2$
Calorimeter 2nd layer	0.025×0.025	$\|\eta\| < 1.40$	0.050×0.025	$1.375 < \|\eta\| < 1.425$
	0.075×0.025	$1.40 < \|\eta\| < 1.475$	0.025×0.025	$1.425 < \|\eta\| < 2.5$
			0.1×0.1	$2.5 < \|\eta\| < 3.2$
Calorimeter 3rd layer	0.050×0.025	$\|\eta\| < 1.35$	0.050×0.025	$1.5 < \|\eta\| < 2.5$
Number of readout channels				
Presampler	7808		1536 (both sides)	
Calorimeter	101 760		62 208 (both sides)	
Liquid Argon (LAr) hadronic endcap				
$\|\eta\|$ coverage			$1.5 < \|\eta\| < 3.2$	
Number of layers			4	
Granularity $\Delta\eta \times \Delta\phi$			0.1×0.1	$1.5 < \|\eta\| < 2.5$
			0.2×0.2	$2.5 < \|\eta\| < 3.2$
Readout channels			5632 (both sides)	
LAr forward calorimeter				
$\|\eta\|$ coverage			$3.1 < \|\eta\| < 4.9$	
Number of layers			3	
Granularity $\Delta x \times \Delta y$ (cm)			LAr Forward Calorimeter (FCal)1: 3.0×2.6	$3.15 < \|\eta\| < 4.30$
			FCal1: ~ four times finer	$3.10 < \|\eta\| < 3.15$, $4.30 < \|\eta\| < 4.83$
			FCal2: 3.3×4.2	$3.24 < \|\eta\| < 4.50$
			FCal2: ~ four times finer	$3.20 < \|\eta\| < 3.24$, $4.50 < \|\eta\| < 4.81$
			FCal3: 5.4×4.7	$3.32 < \|\eta\| < 4.60$
			FCal3: ~ four times finer	$3.29 < \|\eta\| < 3.32$, $4.60 < \|\eta\| < 4.75$
Readout channels			3524 (both sides)	
Scintillator tile calorimeter				
	Barrel		Extended barrel	
$\|\eta\|$ coverage	$\|\eta\| < 1.0$		$0.8 < \|\eta\| < 1.7$	
Number of layers	3		3	
Granularity $\Delta\eta \times \Delta\phi$	0.1×0.1		0.1×0.1	
Last layer	0.2×0.1		0.2×0.1	
Readout channels	5760		4092 (both sides)	

Figure 3.10: Table presenting the number of layers in each sub-calorimeters with the corresponding $|\eta|$ coverage and cell granularity [24].

3.2.5 Muon spectrometer

The muon spectrometer, that will only be briefly described in this section, corresponds to the outer layer of the ATLAS detector. It allows to identify and to measure the momentum of the muons that do not stop in the calorimeter. The trajectory of the muons are bent by the magnetic field produced by the toroidal magnets described in section 3.2.2. The muon spectrometer is shown in Figure 3.11 [24].

It is composed of several cylindrical stations using multi layers of Monitored Drift Tubes (MDTs) and Resistive Plate Chambers (RPCs) in the barrel, of wheels in each end-cap using MDTs and Thin Gap Chambers (TGCs) and of Micromegas in the New Small Wheel (NSW).

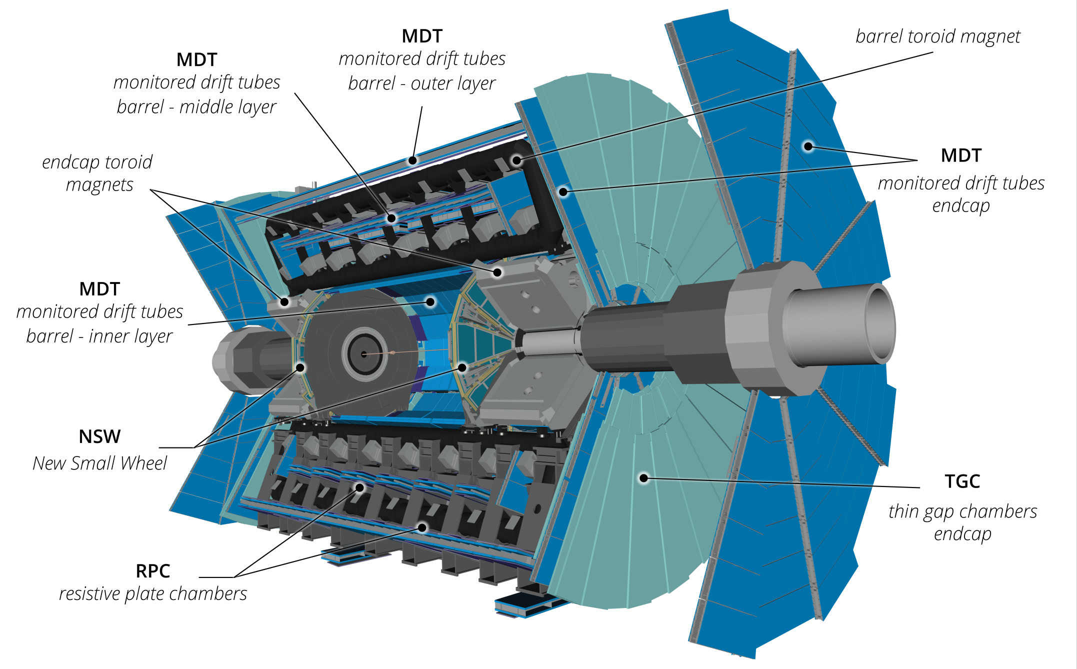


Figure 3.11: Overview of the muon spectrometer [24].

3.3 Data taking with the ATLAS detector

3.3.1 Luminosity

The amount of data delivered by the LHC is quantified by a quantity that is called the instantaneous luminosity $\mathcal{L}$. It is defined such as for a given process, the rate of events observed during a period dN_{event}/dt is equal to the product between the luminosity and the cross-section of the process σ :

$$\frac{dN_{\text{event}}}{dt} = \mathcal{L} \times \sigma \implies \mathcal{L} = \frac{1}{\sigma} \times \frac{dN_{\text{event}}}{dt}. \tag{3.1}$$

The unity of the luminosity is therefore $\text{cm}^{-2}.\text{s}^{-1}$. Given this expression, it becomes important to maximize the luminosity to search for rare processes.

The luminosity at the LHC, which depends on the beam parameters, can be expressed as :

$$\mathcal{L} = \frac{N^2 f N_b}{4\pi\sigma_x\sigma_y} \text{ [28]} \tag{3.2}$$

with N the number of protons per bunch, f the revolution frequency of the bunch, N_b the number of bunches in a beam, and σ_x (σ_y) the bunch size at the interaction point along the x (y) direction. This expression is obtained by supposing that the protons inside the bunches are distributed according to Gaussian distributions along the x, y and z axis. In 2024 during Run-3, the luminosity reached a peak of 2.33×10^{34} $\text{cm}^{-2}\text{s}^{-1}$ [29].

A related quantity is the integrated luminosity $\mathcal{L}_{\text{int}}$, which simply corresponds to the integral of the luminosity with respect to the time :

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \tag{3.3}$$

The integrated luminosity at the LHC is usually expressed using inverse femto-barn fb^{-1} (with the equivalence 1 b = 10^{-24} cm^2). For a process having a certain cross-section σ, the total number of produced events is simply given by :

$$N_{\text{event}} = \sigma \times \mathcal{L}_{\text{int}} \tag{3.4}$$

The cumulative integrated luminosity delivered to ATLAS as a function of time for years between 2011 and 2024 is shown is Figure 3.12 [29]. In total, 28 fb^{-1} and 156 fb^{-1} have been delivered during Run-1 and Run-2 respectively. Since the beginning of Run-3 in 2022, 195 fb^{-1} has already been delivered, with 125 fb^{-1} in 2024 only.

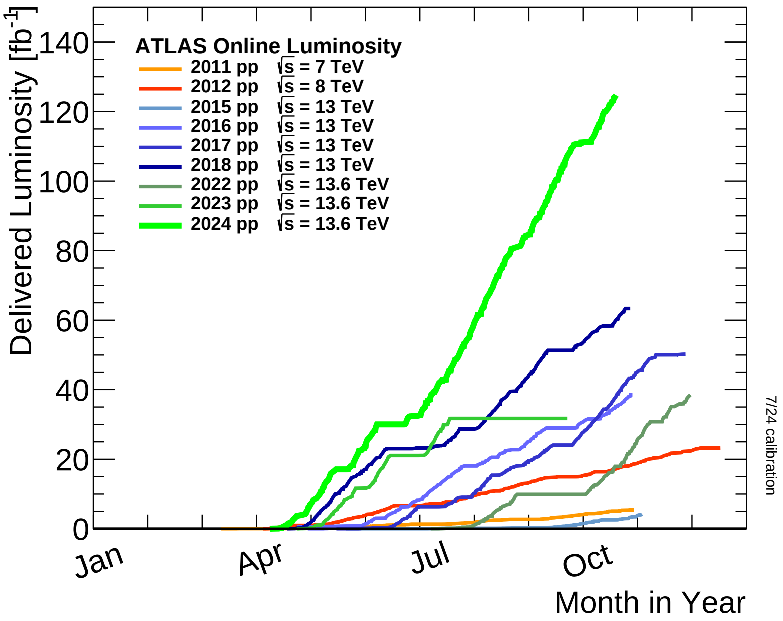


Figure 3.12: Cumulative integrated luminosity delivered to ATLAS as a function of time for years between 2011 and 2024 [29].

Nevertheless, the luminosity effectively recorded by ATLAS is inferior to the one that is delivered by the LHC as shown in Figure 3.13 [29]. The recorded luminosity may reflect for instance the fact that once the LHC announces that stable beams are circulating in the ring, there is a delay before the inner detector is activated and operational. Moreover, the events recorded by ATLAS during periods where some sub-components of the detector were not normally functioning are not used in physic analysis. It is also shown in Figure 3.13, where the label «Good for Physics» corresponds to data that can be effectively used in analysis.

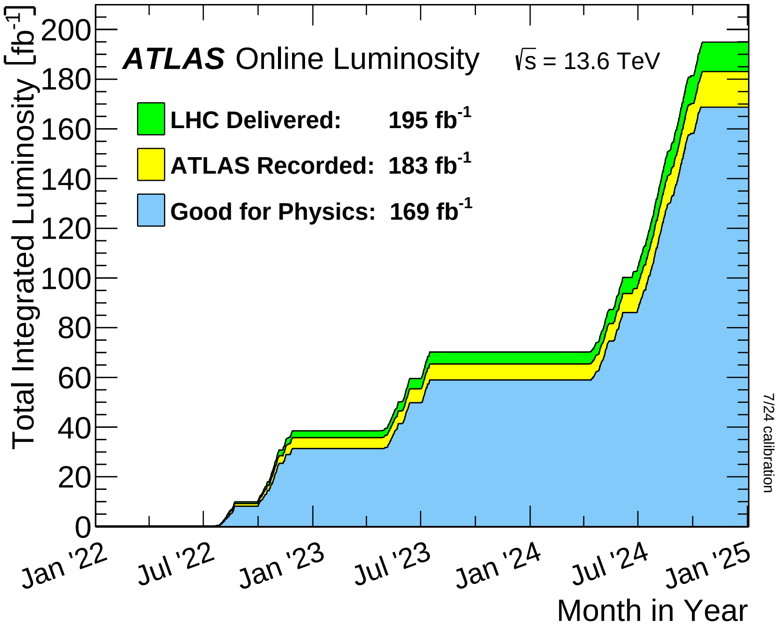


Figure 3.13: Cumulative integrated luminosity delivered to ATLAS as a function of time during Run-3, in addition to the one effectively recorded by ATLAS and the one corresponding to exploitable data [29].

3.3.2 Pile-up

While the high frequency of collision and the important number of protons inside each bunch allow to reach high luminosity values, they also bring some drawbacks such as pile-up.

Indeed, during a single bunch crossing, multiple *pp* interactions occur at the same time. If a hard scattering process occurs in an event, it will be contaminated by additional interactions that may consist in soft and inelastic scatterings : it corresponds to in-time pile-up. Moreover, the time response of some sub-components, such as the LAr calorimeters, can be higher than the time interval that separates two successive bunch crossings. As a consequence, the previous and next bunch crossings can also contaminate a given event : it is named out-of-time pile-up. During the reconstruction of physics objects like the jets, several techniques are used to mitigate the pile-up contribution (they will be detailed in sections 4.3.5 and 4.4.2).

The amount of pile-up can be quantified by considering the time averaged number of interactions per bunch crossing μ. The distribution of μ in the data recorded by ATLAS during the three different Runs is shown in Figure 3.14 [29]. It can be observed that during Run-3, the recorded data are characterized by a larger amount of pile-up compared to Run-1 and Run-2 data.

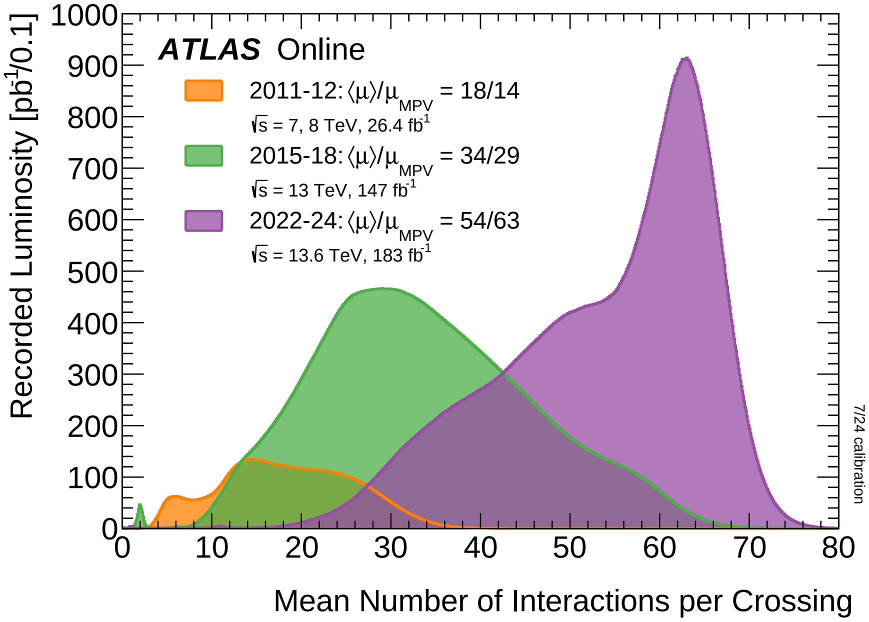


Figure 3.14: Distribution of μ in the data recorded by ATLAS. $< \mu >$ corresponds to the mean of the distribution and μ_{MPV} is the most probable value [29].

3.3.3 Trigger

Given the rate of bunch crossing of 40 MHz and the data size needed to store the information from a collision, it is impossible to keep a record of every single event. Moreover, recording the information from every collision would be unneeded because most *pp* collisions correspond to soft parton scatterings that do not provide any substantial new physics information. As a consequence, a trigger system is used to store the data at a rate that is achievable by identifying and recording potentially interesting events for physics analysis. This trigger system is made of two sub-systems : the Level 1 trigger (L1) and the High Level Trigger (HLT).

The L1 trigger is a hardware based system that combines electronic signals from the calorimeters and the muon spectrometer. It aims to identify interesting collision events within 2.5 μs of the bunch crossing, i.e events presenting large missing transverse energy or high energy leptons, photons or jets candidates. The L1 trigger also defines regions of interest, i.e regions in the detector containing high energy deposits. It allows to reduce the rate from 40 MHz to 100 kHz.

The events that are selected by the L1 trigger are then treated by the HLT, which is a software based system using as inputs data located in the regions of interest identified by the L1 trigger. The HLT is able to perform fast reconstruction of physic objects such as leptons, photons or jets in a faster but similar way as the offline reconstruction that will be detailed in section 4. In this way, it can reconstruct events with higher precision than the L1 trigger and select them according to more complex signatures. The HLT has therefore a processing time of about a few hundreds ms. It allows to reduce the selection rate to 3 kHz. The selected events are then stored in CERN data centers before being processed by the reconstruction software.

3.4 Monte Carlo simulations

The Monte Carlo (MC) simulations consist in reproducing *pp* collisions, simulating the production of particles in each collision before emulating their interactions with the detector. They are used in physic analysis to simulate the events from a theoretical BSM process for instance, to reproduce SM events or to optimize the event selections. They are also notably used in ATLAS to derive the jet calibration (discussed in section 4.4.3) or to evaluate the performances of the algorithms that are used to reconstruct tracks, vertices, jets, etc.

The simulation of the *pp* collisions at a given collision energy and the resulting production of particles is accomplished with particle generators such as Pythia [30], Sherpa [31], Herwig [32] or MadGraph [33]. These generators, or combinations of them, are also able to simulate the processes of parton shower, hadronization, underlying event, ISR and FSR in order to produce simulated events as similar as possible to events obtained in collisions with real data. To achieve this, models with tuned parameters are used to simulate these processes in addition to the use of parton distribution function sets. Finally, the generators can reproduce the decay chains of the unstable particles and then return the final stable particles, including their production vertices and four-momenta. The particle generators are usually employed to simulate a single *pp* interaction per event corresponding to the hard scattering process. In ATLAS, pile-up interactions are generally simulated separately and adjusted to match the observed pile-up conditions at the LHC, before being added to the hard scattering process.

To simulate the interaction of the stable particles with the detector, ATLAS uses the Geant4 software [34]. This software allows to model precisely the ATLAS detector, including both the active materials where the energy deposits are measured and the inactive components (cables, supporting construction) which can absorb particles. Geant4 allows to simulate the progression in the detector of the stable particles and their interaction with the magnetic fields and detector material, producing a list of energy deposits in the different sub-detectors. Finally, all the simulated energy deposits are converted into digitized signals which have a similar format as the signals measured by the ATLAS detector with real data. A trigger simulation providing trigger decisions similar to the ones available for data is also used. In the end, the digital signals are used to reconstruct physical objects by using the same algorithms as the ones that are applied to signals from data.

Finally, the MC simulations allow to provide information on the generated particles before their interaction with the detector. These particles are qualified as «truth» particles and can be used to evaluate track or vertex reconstruction efficiencies for instance.

Chapter 4

Reconstruction of physical objects with the ATLAS detector

Contents

In ATLAS, the energy deposits in the calorimeters cells and the hits in the inner detector and the muon spectrometer are used to reconstruct physic objects that can be used for analyses. These objects can be tracks, vertices, jets, photons and leptons. Only the reconstruction of the first three will be detailed in this section because they are the main objects that are considered in the analysis described in section 5.

4.1 Track reconstruction

The challenge of the track reconstruction algorithm is to identify and combine hits in the tracking detectors that are originating from a unique charged particle in order to build a track corresponding to the particle trajectory. The track is characterized by five parameters (d_0, z_0, ϕ, θ, q/p) that are estimated by the algorithm : d_0 and z_0 are the transverse and longitudinal impact parameters of the track, ϕ and θ are the azimutal and polar angle and q/p is the charge of the particle divided by the momentum (cf Figure 4.1 [35]).

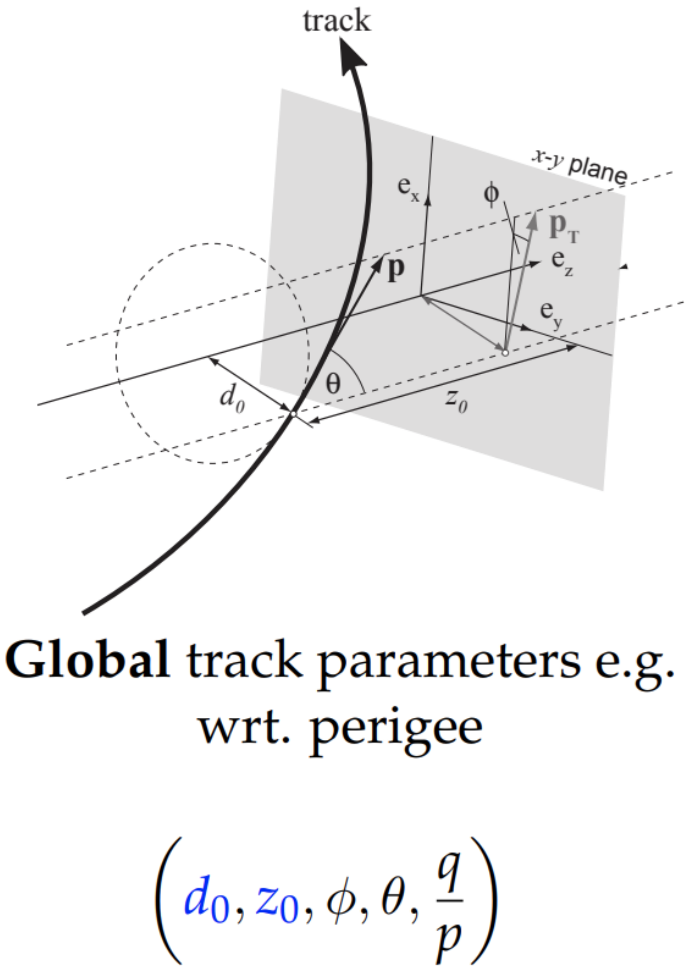


Figure 4.1: Diagram presenting the different track parameters [35].

The reconstruction of tracks originating from charged particles produced near the pp interaction point is divided in two different steps that are executed consecutively [36]. First, the reconstruction process is run from the inside to the outside of the inner detector, i.e it starts from hits in the Pixel and SCT sub-detectors («Primary Tracking»). Then, a second run is performed from the outside to the inside, starting from leftover hits in the TRT («Back Tracking»). These two steps are schematized in Figure 4.2 [36]. A third step named «Large Radius Tracking» is also performed to reconstruct tracks that are highly displaced from the interaction point (i.e at large d_0).

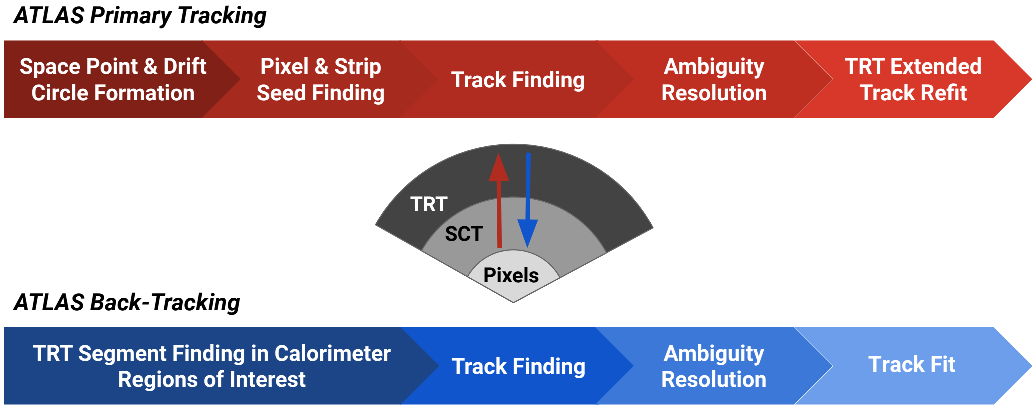


Figure 4.2: Overview of the primary tracking and back tracking chains used in the track reconstruction [36].

Primary Tracking The primary tracking algorithm starts by forming track seeds made of three space-points in either the Pixel or the SCT sub-detectors or a combination of the two. The hits are combined if they are likely to originate from a single charged particle produced close to the interaction point. Then, based on the estimated trajectory of a given seed, search roads are built through the detector in order to extend the seed with remaining hits thanks to a combinatorial Kalman Filter [37] to produce a silicon track candidate.

In order to resolve overlaps between track candidates and to reject fake tracks, an ambiguity resolution algorithm is executed. It ranks every track candidate according to quality criteria, such as the number of hits shared with other candidates : the algorithm rejects lower-quality candidates that share a large number of hits with higher-quality ones. The track candidates that pass the ambiguity resolution step are then fitted using a global χ^2 method in order to obtain the final track parameters.

Finally, an extension of the track to the TRT sub-detector is executed by collecting the drift tubes that are consistent with the extrapolated trajectory of the track. A refit of the entire track that include these TRT drift-circles is performed and allows to improve the estimation of the track parameters.

Back Tracking The back tracking step allows to increase the track reconstruction acceptance for particles that are produced at higher distance from the interaction point, such as electrons coming from photon conversions that happen in the dense area of the detector. This secondary reconstruction uses the remaining hits after the primary step has been performed. In addition, this track reconstruction is only performed in regions of the inner detector that are consistent with energy deposits in the electromagnetic calorimeter.

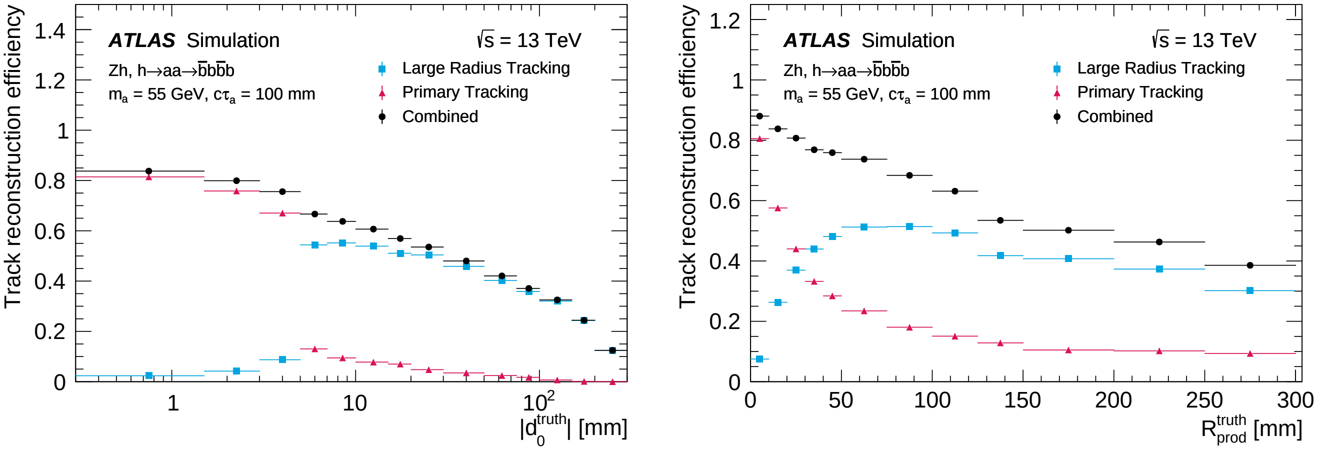
The algorithm starts by considering compatible hits in the TRT that form what is called segments. For each of these, track seeds made of two space-points in the SCT close to the TRT are constructed, before being extended through the inside of the SCT and Pixel sub-detectors with the same combinatorial Kalman Filter used in the primary tracking step. An ambiguity resolution algorithm is then run to filter the track candidates before the final tracks (that include segments from TRT) are fitted and their parameters estimated.

Large Radius Tracking (LRT) In order to reconstruct tracks with a large transverse impact parameter d_0 coming from the decay of long-lived particles (LLPs), an additional track reconstruction pass through the leftover hits is required. This pass is similar to the primary tracking step, the main difference being that it is configured with looser requirements on tracks impact parameters (cf Figure 4.3 [38]).

<table>
<tr><th>Selection criteria</th><th>Primary (<i>inside-out</i>)</th><th>LRT</th></tr>
<tr><td>max. $|d_0|$ [mm]</td><td>5</td><td>300</td></tr>
<tr><td>max. $|z_0|$ [mm]</td><td>200</td><td>500</td></tr>
<tr><td>min. p_{T} [GeV]</td><td>0.5</td><td>1</td></tr>
<tr><td>max. $|\eta|$</td><td>2.7</td><td>3.0</td></tr>
<tr><td>max. silicon holes</td><td>2</td><td>1</td></tr>
<tr><td>max. double holes</td><td>1</td><td>0</td></tr>
<tr><td>max. holes gap</td><td>2</td><td>1</td></tr>
<tr><td>road width [mm]</td><td>12</td><td>5</td></tr>
<tr><td>seeding</td><td>Pixels and SCT</td><td>SCT only</td></tr>
<tr><td>max. seeds per middle Pixel SP</td><td>1</td><td>–</td></tr>
<tr><td>max. seeds per middle SCT SP</td><td>5</td><td>1</td></tr>
<tr><th colspan="3">Common selection criteria</th></tr>
<tr><td>min. silicon hits</td><td colspan="2">8</td></tr>
<tr><td>min. unshared silicon hits</td><td colspan="2">6</td></tr>
<tr><td>max. track χ^2/n_{DoF}</td><td colspan="2">9</td></tr>
<tr><td>keep all confirmed seeds</td><td colspan="2">true</td></tr>
</table>

Figure 4.3: Table presenting the selection criteria used for primary and large radius tracking algorithms (a hole corresponds to a missing hit on an active module where one was expected based on the particle trajectory) [38].

Studies to evaluate the performance of the large radius tracking reconstruction have been performed. They consist in examining the track reconstruction efficiency for charged particles coming from the decay of BSM LLPs. This efficiency is evaluated in MC simulated events and is defined as the proportion of truth particle that are matched to a reconstructed track. In Figure 4.4 [38], the track reconstruction efficiency is shown as a function of the transverse parameter of the truth particle $|d_0^{\text{truth}}|$ and as a function of its radius of production $R_{\text{prod}}^{\text{truth}}$ for both primary and large radius tracking, either separately or combined. It can be observed that the large radius tracking allows to significantly improve the reconstruction of tracks for charged particles produced at large radius i.e at $R_{\text{prod}}^{\text{truth}} > 50$ mm and with $|d_0^{\text{truth}}| > 5$ mm, while the primary tracking is highly efficient to reconstruct charged particles produced close to the interaction point.



(a) (b)

Figure 4.4: Primary tracking, large radius tracking and combined track reconstruction efficiency for charged particles produced by the decay of BSM LLPs as a function of the truth particle (a) transverse impact parameter $|d_0^{\text{truth}}|$ and (b) radius of production $R_{\text{prod}}^{\text{truth}}$ [38].

4.2 Vertex reconstruction

Based on the collection of reconstructed tracks, two type of vertices can be reconstructed : primary vertices which are the locations of *pp* collisions, and secondary vertices associated to the decay of LLPs that have been produced at primary vertices. Primary and secondary vertex reconstructions are made with two different algorithms.

4.2.1 Primary vertex

The algorithm used by ATLAS during Run-3 to reconstruct primary vertices is named « adaptative multi-vertex finder and fitter » (AMVF) [39]. It is an iterative algorithm that reconstruct one vertex at a time and that is divided into several steps.

First, the reconstructed tracks that pass quality criteria and that are not yet assigned to any vertex candidate are used in order to determine the most likely position of a new primary vertex (named a seed) along the beam axis. This is done by considering the track density repartition along the beam axis.

Once a seed has been found and its position has been computed, all the nearby tracks are selected and associated to the seed in order to create a vertex candidate. These can be unassigned tracks or tracks already assigned to one or more previously found vertex candidates. In this way, each track is likely to be associated to multiple vertex candidates, and the compatibility between a track and an assigned vertex is quantified with a weight.

The parameters of the assigned tracks are then used to fit the position of the vertex candidate. Moreover, each time a new vertex candidate is fitted, all other vertex candidates that are connected to it through a chain of common tracks are also refitted at the same time. The track weights are progressively adjusted based on their compatibility, and all the linked vertices are refitted after each adjustment.

Finally, the vertex candidate has to satisfy some criteria before being accepted. It must be composed of at least two compatible tracks, and it cannot be made only of tracks that are also compatible with a vertex that has been previously determined during the algorithm. Moreover, the fit vertex position cannot be within 3σ of any other candidates, based on the computed fit errors on the positions of both candidates. Furthermore, with w designating the weight of the tracks assigned to the vertex candidate, $\sum w^2/\sum w$ must be higher than 2/3.

If the vertex candidate is accepted, the compatible tracks will not be used for the seed-finding step during subsequent iterations while if it is rejected, it is only the most compatible tracks that are removed. The algorithm stops either when less than two tracks remain for the seed-finding step or when no seed can be determined.

Among all the accepted vertices, the one with the highest $\sum p_T^2$, with the sum occurring on its associated tracks, is classified as the hard-scattering primary vertex of the event and the other ones are designated as pile-up vertices.

4.2.2 Displaced secondary vertex

A different algorithm is used to reconstruct secondary vertices coming from the decays of LLPs [40]. The reconstruction of such vertices is essential for the analysis presented in section 5.

All the reconstructed tracks, i.e either standard or large-radius tracks, are used during the reconstruction. Some pre-selections are applied on them : their transverse momentum must satisfy $p_T > 1$ GeV and they must not be associated to any primary vertex, either the hard-scattering or pile-up ones. Additional requirements are made on the number of hits in the different sub-detectors of the inner detector in order to reduce contribution from low-quality tracks.

The vertex reconstruction starts by forming all possible pairs of the preselected tracks satisfying $|d_0| > 2$ mm. For each pair, the position of the corresponding vertex is estimated based on the track parameters. Then, if both tracks have small impact parameters with the estimated vertex position as reference, a fit is performed to extrapolate the tracks : the vertex candidate is then either accepted or rejected given the quality of the fit. The two tracks composing a given vertex candidate have then to satisfied a «hit-pattern consistency». It consists in requiring that both tracks should have hits in the layers of the tracker at higher radius than the vertex position, while they should not have hits in layers with smaller radius than the vertex position as shown in Figure 4.5 [40].

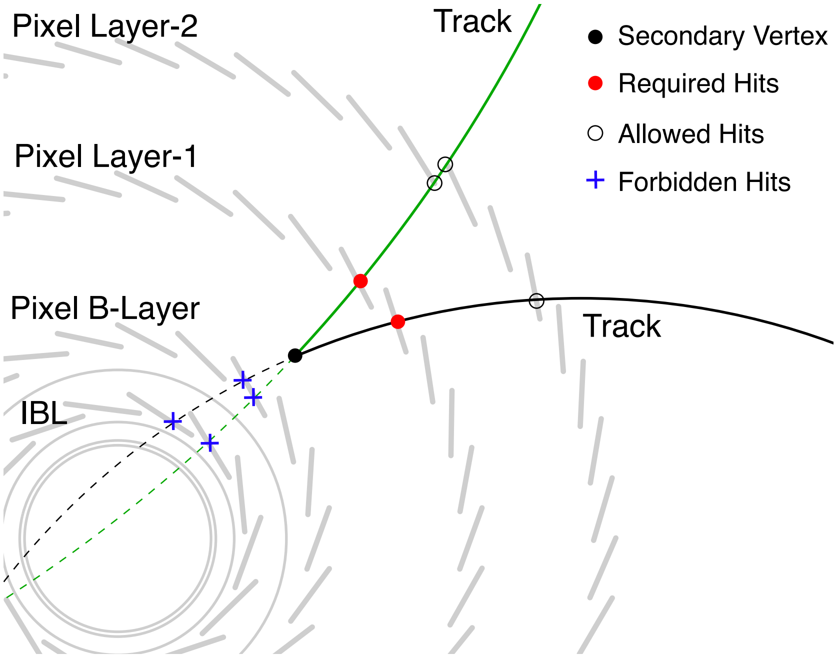


Figure 4.5: Transverse view of the pixel detector showing for a two-track secondary vertex candidate the hits that are either required, allowed or forbidden in the different layers in order to fulfil the hit-pattern consistency [40].

The next step is to form multi-track vertices. Indeed, at this stage, tracks originating from the decay of a single LLP can be used to form several two-tracks vertices. It is also possible that a single track is used in multiple two-track vertex candidates. From all the paired tracks, a compatibility graph can be built by representing each track as a node and each two-track vertex as an edge between the two corresponding nodes. Then, groups of fully compatible tracks (i.e with corresponding nodes all connected to each other in the compatibility graph) can be extracted and fitted as single multi-track vertices.

Each track is now associated to only one secondary vertex candidate. However, the algorithm still allows to reconstruct multiple vertices corresponding to a single long-lived particles decay. The next step is then to merge nearby vertices into a single one. The algorithm aims to merge each vertex candidate by first considering the ones with the lowest track multiplicities, and trying to combine them with the candidates with the highest ones. Two vertices are merged if the difference between the two candidate positions is lower than 10σ, with σ corresponding to the uncertainty on the difference between the two vertex positions. Additional tests (detailed in [40]) are performed to determine whether the vertices should be merged or not. Finally, any remaining vertices that are within 1 mm are forced to merge, and all the obtained merged vertices are refitted with all the tracks coming from both input vertices.

The last step of the reconstruction procedure is to associate the reconstructed tracks that do not pass the selection made during the track-pair forming step, including tracks associated to primary vertices. Indeed, this association can allow to improve the reconstruction of vertex kinematic variables. A track is associated to a vertex if its impact parameters computed with the vertex position as reference are sufficiently small. In the case of a track associated to more than one secondary vertex, the track will be attached to the vertex with the highest-track multiplicity. A final fit is then performed for each secondary vertex including all the attached tracks. The invariant mass of each vertex can be computed from the sum of the attached tracks four-vectors, assuming a charged pion mass for the tracks and using the track momentum vectors at the position of the vertex.

In Figure 4.6 [40], the total vertex reconstruction efficiency (i.e encapsulating the track reconstruction and selection efficiencies and the pure vertex reconstruction efficiency) for different BSM processes involving LLPs decaying into charged particles is shown as a function of the decay radius of the truth LLPs. This efficiency is evaluated as the fraction of individual LLP decays that are matched to a reconstructed secondary vertex with the matching procedure being described in [40].

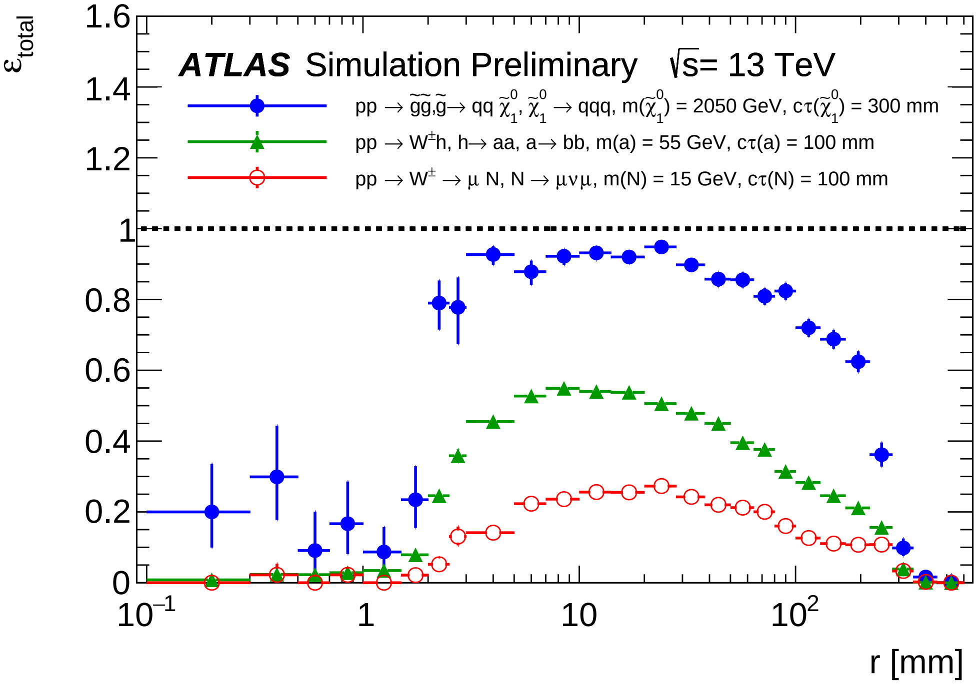


Figure 4.6: Secondary vertex reconstruction efficiency as a function of the BSM LLPs decay radius for three different processes [40].

It can be seen that the efficiency starts being important above $r \sim 2$ mm, which is related to the $|d_0| > 2$ mm cut applied to the tracks used to build two-track vertex candidates. The efficiency decreases with increasing r, reflecting the fact that the large radius tracking efficiency degrades with increasing radius, leading to a smaller number of reconstructed tracks and less possibilities to form vertex candidates at large r. Furthermore, for a given lifetime, the average LLP boost increases with r, inducing more collimated decay products and a more complicated vertex topology to reconstruct. Moreover, the efficiency at a fixed decay radius increases with the vertex track multiplicity, the latter being correlated to the mass of the LLP and to the number and type (quark or lepton) of the charged particles coming from the LLP decay. This explains the differences observed in terms of reconstruction efficiency between the different LLPs that have been considered.

4.3 Jet constituent reconstruction

To reconstruct jets, the first step is reconstructing the jet constituents. These constituents can be built either only from calorimeter signals or from a combinaison between inner detector and calorimeter information.

4.3.1 Topological clusters

Jet constituents that are reconstructed only from calorimeter signals are named «topological clusters» (topo-clusters) [41]. A topo-cluster corresponds to a group of adjacent individual calorimeter cells designed to extract significant signals coming from particle energy depositions in the calorimeters while removing contribution induced by noise. A single topo-cluster can contain the energy contribution from several particles, and the energy deposition of an individual particle can be split into distinct topo-clusters.

The gathering of calorimeter cells is determined with the use of the cell signal significance $\varsigma_{\text{cell}}^{\text{EM}}$ defined as :

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

with $E_{\text{cell}}^{\text{EM}}$ representing the cell signal and $\sigma_{\text{noise,cell}}^{\text{EM}} = \sqrt{(\sigma_{\text{noise,cell}}^{\text{electronic}})^2 + (\sigma_{\text{noise,cell}}^{\text{pile-up}})^2}$ the expected noise in the cell coming both from electronic and pile-up. EM means that both $E_{\text{cell}}^{\text{EM}}$ and $\sigma_{\text{noise,cell}}^{\text{EM}}$ are measured on the electromagnetic energy scale at which the energy calibration is such that the energy deposited in the electromagnetic calorimeter by electrons and photons is correctly evaluated. At this scale, corrections taking into account the possible loss of energy for hadrons during hadronic showers (through the absorption of low energy hadrons in passive regions of the calorimeters for instance) are not included.

To form topo-clusters, the algorithm starts by selecting cells passing the criterion :

$$|\varsigma_{\text{cell}}^{\text{EM}}| > S \text{ with } S = 4 \tag{4.2}$$

and by ordering them in decreasing $\varsigma_{\text{cell}}^{\text{EM}}$. Each of these seed cells forms a proto-cluster and is extended with neighbor cells that satisfy either :

$$|\varsigma_{\text{cell}}^{\text{EM}}| > N \text{ with } N = 2 \tag{4.3}$$

or :

$$|\varsigma_{\text{cell}}^{\text{EM}}| > P \text{ with } P = 0 \tag{4.4}$$

Neighbor cells as defined as two calorimeter cells either directly adjacent in a given sampling layer or being in adjacent layers and having at least a partial overlap in the (η, ϕ) plane. If a neighbor cell satisfies the threshold defined by N, its neighbor are also included to the proto-cluster. If a neighbor cell is a proto-cluster satisfying the threshold defined by S, the two proto-clusters are merged. Moreover, if a cell is adjacent to two different proto-clusters and passes the threshold defined by N, the two proto-clusters are merged as well. Finally, this procedure of extending a proto-cluster ends when the last set of neighbor cells pass the threshold defined by P but not the one defined by N, and the algorithm then starts to consider the next seed.

In this way, topo-clusters are formed as a core of highly significant cells surrounded by less significant ones. Their energy and barycentre coordinates are calculated as $E_{\text{clus}}^{\text{EM}} = \sum E_{\text{cell}}^{\text{EM}}$, $\eta_{\text{clus}} = (\sum |E_{\text{cell}}^{\text{EM}}| \cdot \eta_{\text{cell}})/(\sum |E_{\text{cell}}^{\text{EM}}|)$ and $\phi_{\text{clus}} = (\sum |E_{\text{cell}}^{\text{EM}}| \cdot \phi_{\text{cell}})/(\sum |E_{\text{cell}}^{\text{EM}}|)$ with the sum occurring on all the associated cells. Given the fact that each topo-cluster is interpreted as a massless pseudo-particle, $E_{\text{clus}}^{\text{EM}}$, η_{clus} and ϕ_{clus} are sufficient to determine its four-vector.

It can be noted that the energy of a cell $E_{\text{cell}}^{\text{EM}}$ can be negative, as the result of fluctuations mainly caused by out-of-time pile-up. Indeed, the pulse shape in the cells of the electromagnetic calorimeter contains a negative part at response time higher than 100 ns as shown in Figure 4.7 [41] : this shape has been designed in this manner in order to cancel in average contributions from in-time pile-up by the ones from out-of-time pile-up coming from previous bunch crossings.

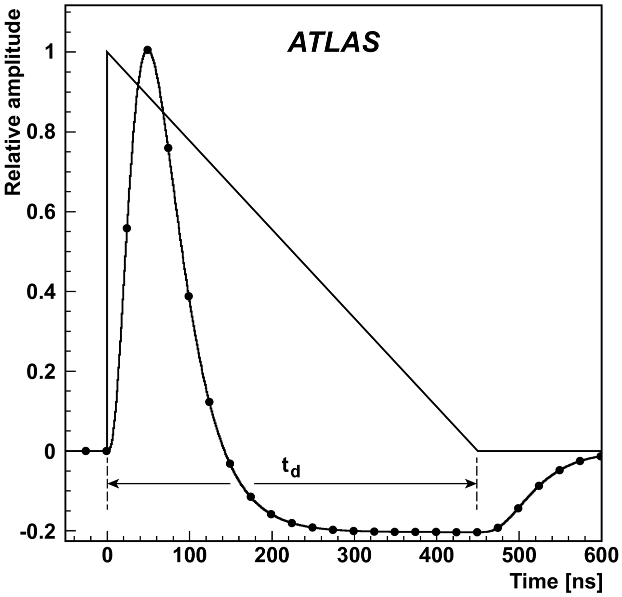


Figure 4.7: Pulse shape in the LAr calorimeters with t_d being the charge collection time [41].

Because the thresholds are applied to the absolute value of $\varsigma_{\text{cell}}^{\text{EM}}$, cells with large negative signals can be used as seeds to form topo-clusters and cells with negative signals can also be considered during the extension step of a proto-cluster having a positive energy. Topo-clusters reconstructed from negative seed cells will often have a total energy $E_{\text{clus}}^{\text{EM}}$ negative, because of the correlation between the seed cell and its neighbors that are also likely to have a negative signal. The number of such negative clusters can be used as an estimator of out-of-time pile-up induced by particle depositing energy more than 100 ns before the considered event. On the other hand, topo-clusters coming from energy deposits in previous bunch crossings but closer in time (i.e within a time interval inferior to 100 ns) are likely to have a positive energy. Topo-clusters having $E_{\text{clus}}^{\text{EM}} < 0$ can be used to determine an average cancellation of topo-clusters from out-of-time pile-up having $E_{\text{clus}}^{\text{EM}} > 0$.

Finally, only topo-clusters with a strictly positive energy are considered to serve as jet constituents, because negative topo-clusters are considered to be uncorrelated to the reconstructed signal of the considered event.

An additional step named «Local hadronic calibration» can be applied on each topo-cluster to correct the non-compensating calorimeter response resulting in smaller signals for hadrons compared to electrons or photons depositing the same energy. It also allows to remedy the losses of signals either due to clustering (coming from the intrinsic noise suppression that can remove true signal in particular regions of the detector) or energy deposits in inactive material.

The corrections applied depend on the particles at the origin of the energy deposit forming the topo-cluster, i.e either hadronic or electromagnetic. The first step of the calibration is to compute for each topo-cluster the probability $\mathcal{P}_{\text{clus}}^{\text{EM}}$ that it has been generated by an electromagnetic shower. Such probabilities have been determined using MC simulations of single charged and neutral pions entering the calorimeters, producing respectively hadronic and electromagnetic showers. The probabilities have been derived in bins defined by topo-cluster variables such as $E_{\text{clus}}^{\text{EM}}$, η_{clus} and other ones describing the shape and the internal signal distribution of the topo-clusters.

Once $\mathcal{P}_{\text{clus}}^{\text{EM}}$ has been computed for a given cluster, a weight is computed for each cell of the topo-cluster in order to correct the loss of energy during hadronic showers with respect to electromagnetic ones. It is computed as $w_{\text{cell}}^{\text{cal}} = \mathcal{P}_{\text{clus}}^{\text{EM}} \cdot w_{\text{cell}}^{\text{em-cal}} + (1 - \mathcal{P}_{\text{clus}}^{\text{EM}}) \cdot w_{\text{cell}}^{\text{had-cal}}$, where $w_{\text{cell}}^{\text{em-cal}}$ and $w_{\text{cell}}^{\text{had-cal}}$ represent respectively the electromagnetic and hadronic cell correction factors derived with MC simulations of pions as a function of variables describing the cell itself and the topo-cluster it belongs to. Additional weights are applied to the cells in order to correct for signal losses due to clustering and energy deposits in dead material.

Finally, the energy of a topo-cluster after the local calibration is given by $E_{\text{clus}}^{\text{LCW}} = \sum w_{\text{cell}}^{\text{LCW}} \cdot E_{\text{cell}}^{\text{EM}}$ with the sum occurring on all the associated cells and $w_{\text{cell}}^{\text{LCW}}$ taking into account the different corrections. The topo-cluster barycentre coordinates η_{clus} and ϕ_{clus} are re-computed since they may be different from the ones at the EM scale. The topo-clusters obtained after this calibration are said to be at the LCW scale and are named «LCTopo clusters», while the uncalibrated ones are at the EM scale and are designated as «EMTopo clusters».

In order to exploit the complementarity between the calorimeters and the inner detector informations, tracks and topo-clusters can be combined to build new type of jet constituents. In ATLAS, different manners of combining them are used, allowing to construct different objects : the Particle Flow (PFlow), The Track-CaloCluster (TCC) and the Unified Flow Object (UFO).

4.3.2 Particle Flow

The majority of the analyses that were conducted in ATLAS during the Run-1 of the LHC used jets built only from clusters of calorimeter cells. Then, a global jet reconstruction was put in place, performing better as the tracks associated to the jets are also considered.

The Particle Flow constituents [42] have been developed as an alternative to topo-clusters to take advantage of the complementarity between tracks and clusters information. The benefits to this combination are the following :

- The angular resolution of a charged particle is better, as it uses the tracker rather than the calorimeter.
- Tracks are reconstructed only for charged particles while the calorimeter is able to measure energy deposits for both charged and neutral particles.
- Each reconstructed track can be associated to a vertex from which it originates, and thus tracks from the hard scattering vertex can be differentiated from tracks starting at pile-up vertices.
- At high energy, the calorimeter energy resolution is greater than the tracker transverse momentum resolution, while it is the opposite at low energy.
- Low-p_T charged particles that should be part of a jet can move out of the jet cone due to the magnetic field by the time they reach the calorimeters. By considering the corresponding track coordinates, these particles can be included in the jet.
- The tracker acceptance is limited to $|\eta| < 2.5$, while the calorimeters are able to cover up to $|\eta| < 4.9$.

The principle of the Particle Flow algorithm is to determine whether the calorimeter or the inner detector measurement should be used to construct each constituent. If the track of a charged particle is used, the corresponding energy has to be subtracted from the calorimeter energy measurement to avoid double counting of the particle energy.

The Particle Flow algorithm is detailed in Figure 4.8 [42]. It starts with the selection of tracks and topo-clusters at the LCW scale. Tracks are required to have at least nine hits in the silicon detectors, and must have no missing Pixel hits when they would be expected : with these requirements, the number of badly reconstructed tracks is minimized. Moreover, tracks with $p_T > 40$ GeV are not selected because the particles that create such tracks are often poorly isolated, and that could compromise the energy subtraction procedure that is expected to be precise. Tracks that are matched to electrons or muons are also rejected : the algorithm is optimized to subtract energy from hadronic particles.

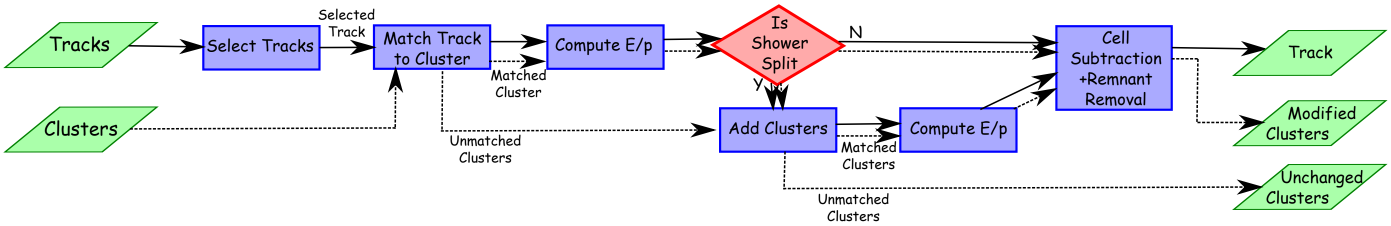


Figure 4.8: Schematic of the Particle Flow algorithm [42].

The algorithm continues by matching each selected track to one topo-cluster. It is done by considering the variable defined as :

$$\Delta R' = \sqrt{(\frac{\Delta\phi}{\sigma_\phi})^2 + (\frac{\Delta\eta}{\sigma_\eta})^2} \tag{4.5}$$

where $\Delta\phi$ and $\Delta\eta$ are the distances between the barycentre of the cluster and the track that is extrapolated to the second layer of the EM calorimeter, and σ_η and σ_ϕ correspond to the angular topo-clusters width [41]. The topo-clusters are preliminarly selected by demanding $E^{\text{clus}}/p^{\text{trk}} > 0.1$, with E^{clus} the topo-cluster energy and p^{trk} the track momentum. Then, the closest pre-selected cluster in $\Delta R'$ to the track is taken to be the matched topo-cluster. The $\Delta R'$ distance between the matched topo-cluster and the track has to be below 1.64 : if no cluster is found in a cone of size $\Delta R' = 1.64$ around the track position, the track is retained and no energy subtraction is performed.

For a track that has been matched to a topo-cluster, the expected deposited energy $\langle E_{\text{dep}}(p^{\text{trk}})\rangle$ that will be subtracted from the topo-cluster has to be evaluated. It can be written as :

$$\langle E_{\text{dep}}(p^{\text{trk}})\rangle = p^{\text{trk}}\langle E^{\text{clus}}_{\text{ref}}/p^{\text{trk}}_{\text{ref}}\rangle \tag{4.6}$$

where $\langle E^{\text{clus}}_{\text{ref}}/p^{\text{trk}}_{\text{ref}}\rangle$ has been determined by considering simulated samples of signal-charged particles with a given $p^{\text{trk}}_{\text{ref}}$, with $E^{\text{clus}}_{\text{ref}}$ computed by summing the energies of the topo-clusters contained in a $\Delta R = 0.4$ cone around the track extrapolated position. $E^{\text{clus}}_{\text{ref}}/p^{\text{trk}}_{\text{ref}}$ distributions are evaluated in bins of p_T and η of the track. One can determine the mean value and the standard deviation $\sigma(E_{\text{dep}})$ of the expected deposited energy distributions.

The next step is to add other topo-clusters to the one that has been firstly matched with the procedure previously described. Indeed, particles do not always deposit their energy into a single topo-cluster, especially for the highest track p_T value. To determine if the hadronic shower is split into several topo-clusters or not, the energy of the matched topo-cluster E^{clus} is compared to the expected energy $\langle E_{\text{dep}}\rangle$. If $E^{\text{clus}} < \langle E_{\text{dep}}\rangle - \sigma(E_{\text{dep}})$, it is decided that other topo-clusters have to be added to the already matched one : the clusters within a cone of $\Delta R < 0.2$ around the track position are matched to the track.

Finally, the energy subtraction is performed. If $\langle E_{\text{dep}} \rangle$ is superior to the total energy of the collection of matched topo-clusters, the latter are simply removed. Otherwise, the subtraction is executed cell-by-cell : the cells of the matched topo-clusters are ranked in descending order of energy, and the energy is removed from the most energetic cell to the lowest one until the total subtracted energy reaches $\langle E_{\text{dep}} \rangle$.

The last step is to consider the remaining energy in the matched topo-clusters after the energy subtraction procedure, and to compare it to $1.5\sigma(E_{\text{dep}})$: if the remnant energy is below $1.5\sigma(E_{\text{dep}})$, the remaining cells are removed because the corresponding energy is compatible with the particle shower and its fluctuations. Otherwise, the remaining topo-clusters are kept, and correspond to a case where they contain an energy contribution from more than one particles.

The process is then repeated for all the tracks of the event. The final set of constituents then corresponds to the selected tracks that are called Charged Particle Flow Objects, the remaining topo-clusters after the track energy subtraction step and the unmodified topo-clusters (i.e unmatched to any tracks) that are both named Neutral Particle Flow Objects. The four-vector of a Charged Particle Flow is computed by considering the track momentum and by setting its mass equal to the charged pion mass. Concerning the topo-clusters modified by the energy cell subtraction, the barycentre coordinates η and ϕ (and thus the four-vector) are recomputed.

4.3.3 TrackCaloCluster

The TrackCaloCluster constituents [43] also correspond to objects that are made by combining inner detector tracks and clusters information, and exploiting both the superior angular resolution of the tracker and the excellent energy resolution of the calorimeters. In this case, the advantages of using these objects is to improve the resolution of the reconstructed jet mass and substructure variables. The latter aim at describing the internal energy structure within a jet and are fundamental in the tagging of jets initiated by bottom or top quarks for instance. They also help to discriminate between jets initiated by light quarks and specific jet topologies, like jets from boosted Z or W bosons decaying to a pair of quarks that will produce two highly collimated jets.

Like the Particle Flow algorithm, the TCC one is also based on a geometric matching between tracks and topo-clusters, but it relies on an energy sharing procedure instead of an energy subtraction algorithm. It is also based on an extrapolation of the tracks to the calorimeters, which is more and more accurate as the tracks momentum increases because the charged particles are less curved by the magnetic field. Moreover, the track extrapolation uncertainty is found to be generally smaller than the angular width of the topo-clusters as defined in the previous section, for a large range of p_T and η track values.

The first step of the algorithm is to match every track passing quality criteria to every topo-cluster calibrated at the LCW scale. In this matching process, the track extrapolation uncertainty is first compared to the width of the topo-cluster : if the extrapolation uncertainty is higher than the topo-cluster width, the track is not considered. Otherwise, the angular distance ΔR between the extrapolated track and the centroid of the topo-cluster is compared to $\sqrt{\sigma^2_{\text{cluster}} + \sigma^2_{\text{track}}}$, where $\sigma_{\text{cluster}} = \sqrt{\sigma^2_\eta + \sigma^2_\phi}$ is the topo-cluster width, and σ_{track} is the track extrapolation uncertainty : if $\Delta R < \sqrt{\sigma^2_{\text{cluster}} + \sigma^2_{\text{track}}}$, the track and the cluster are considered to be matched. In particular at high p_T, this requirement becomes $\Delta R \lesssim \sigma_{\text{cluster}}$.

Three types of TCC objects are built with this matching process : tracks from the hard scattering primary vertex that are matched to topo-clusters, topo-clusters that are not matched to any tracks from any vertex, and tracks from the primary vertex that are not matched to any topo-clusters. They are respectively named combined, neutral and charged TCC constituents. The tracks coming from pile-up vertices that are matched to topo-clusters are not used to build TCC objects.

The next step of the algorithm is to build a new four-vector for each TCC constituent and several cases can be possible.

If a single track coming from the hard scattering vertex is matched to only one topo-cluster such as the case ① in Figure 4.9 [43], the four-vector is computed as follow :

$$\text{TCC}_{\textcircled{1}} = (p_T^{c_1}, \eta^{t_1}, \phi^{t_1}, m^{c_1} = 0) \tag{4.7}$$

where c_i and t_i refer respectively to the topo-cluster and the track in the example ⓘ with i $\in [\![1; 7]\!]$. In this example, the combined object obtains the angular information of the track and the transverse momentum of the matched cluster.

For a topo-cluster that is not matched to any tracks (②), the four-vector is directly the topo-cluster's one :

$$\text{TCC}_{\textcircled{2}} = (p_T^{c_7}, \eta^{c_7}, \phi^{c_7}, m^{c_7} = 0) \tag{4.8}$$

In the same way, the four-vector of a track from the hard scattering vertex that is not matched to any topo-clusters (③) is also directly used to build the TCC :

$$\text{TCC}_{\textcircled{3}} = (p_T^{t_6}, \eta^{t_6}, \phi^{t_6}, m^{t_6} = 0) \tag{4.9}$$

Several tracks from the hard scattering vertex can be matched to the same topo-cluster : in this case, one combined TCC will be created for each track. Each of them will used the track angular coordinates and the energy of the topo-cluster will be shared between all the tracks. If only two tracks are matched to the same topo-cluster (④ and ⑤), the four-vectors will be defined as :

$$\text{TCC}_{\textcircled{4}} = (p_T^{c_2} \frac{p_T^{t_2}}{p_T[\boldsymbol{p}^{t_2} + \boldsymbol{p}^{t_3}]}, \eta^{t_2}, \phi^{t_2}, m^{c_2} \frac{p_T^{t_2}}{p_T[\boldsymbol{p}^{t_2} + \boldsymbol{p}^{t_3}]} = 0) \tag{4.10}$$

$$\text{TCC}_{\textcircled{5}} = (p_T^{c_2} \frac{p_T^{t_3}}{p_T[\boldsymbol{p}^{t_2} + \boldsymbol{p}^{t_3}]}, \eta^{t_3}, \phi^{t_3}, m^{c_2} \frac{p_T^{t_3}}{p_T[\boldsymbol{p}^{t_2} + \boldsymbol{p}^{t_3}]} = 0) \tag{4.11}$$

with $p_T[\boldsymbol{p}^a + \boldsymbol{p}^b]$ referring to the transverse momentum of the sum of the four-vectors $\boldsymbol{p}^a$ and $\boldsymbol{p}^b$.

Cases where a track is matched to multiple topo-clusters that can be matched to other tracks imply a more complex energy sharing computation formula. To construct such constituents, the first step is to collect all the topo-clusters that are matched to a given track. For each of those matched topo-clusters, the energy is divided between all the tracks that match the considering topo-cluster, based on $p_T^{t_i}/p_T[\sum_j \boldsymbol{p}^{t_j}]$, the p_T fraction of a track i relatively to all the matched tracks. The importance of each topo-cluster compared to all of the other topo-clusters that are matched to a track is also evaluated. Finally, the four-vector of the TCC will use the angular coordinates of the track and a transverse momentum and a mass computed with an equation that will not be detailed here, but that can be found in [43]. Such cases are illustrated by ⑥ and ⑦ in Figure 4.9.

It can be noted that for a combined TCC, the mass of the resulting four-vector will be zero if the track is matched to a single topo-cluster, and non-zero if it is matched to more than one topo-cluster. Moreover, during the topo-cluster energy sharing, tracks from all vertices (including pile-up ones) are selected if they are matched to the topo-cluster, but tracks from pile-up are not used to build one TCC object. Indeed, the combined TCCs that are reconstructed must originate from the hard scattering vertex.

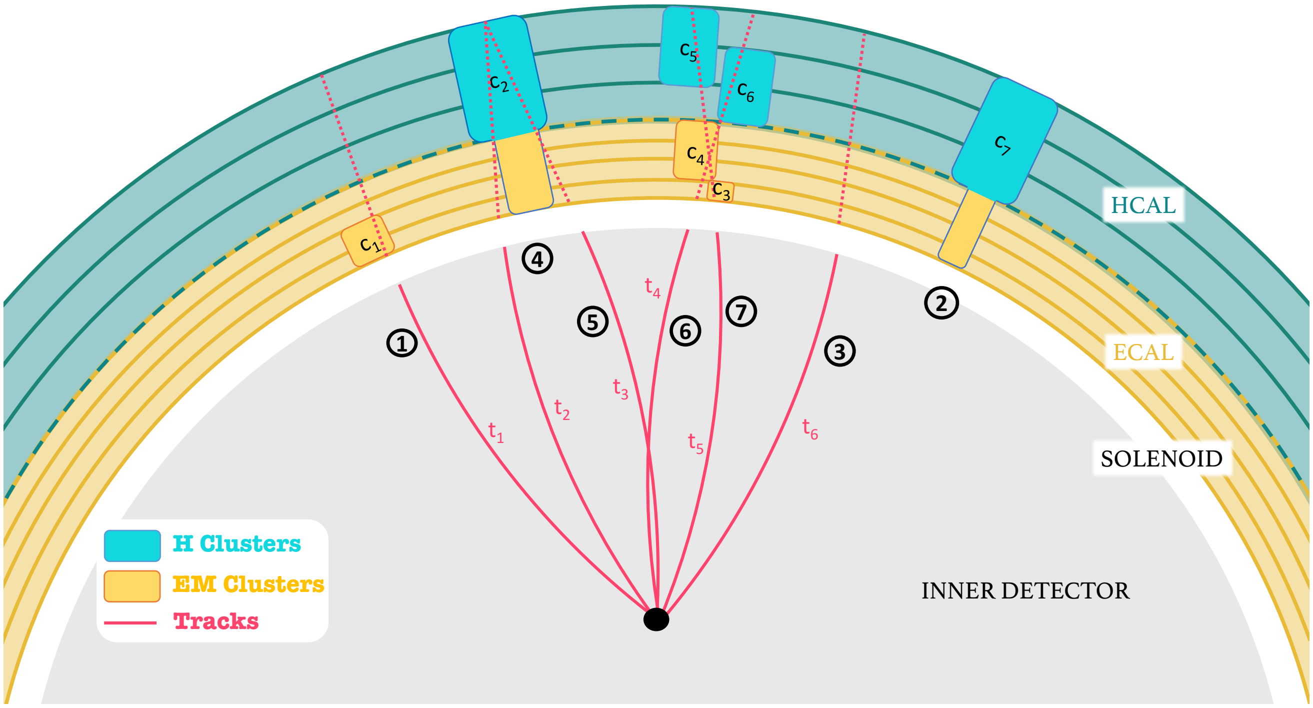


Figure 4.9: Schematic of some possible TCC objects that can be built [43].

4.3.4 Unified Flow Object

The PFlow algorithm has been designed to improve jet reconstruction performances at low p_T, while the TCC algorithm is more efficient at high p_T. Indeed, at higher p_T values, the angular resolution of the inner detector deteriorates, and the PFlow algorithm cannot be used. Moreover, studies have shown that PFlows and TCCs present opposite trends in terms of tagging performance or pile-up stability for instance.

These observations motivated the development of a new type of constituents called Unified Flow Objects (UFOs) [44], that correspond to a combinaison of PFlow and TCC in order to obtain an optimal performance across a large kinematic range. The UFO algorithm is presented in Figure 4.10 [44].

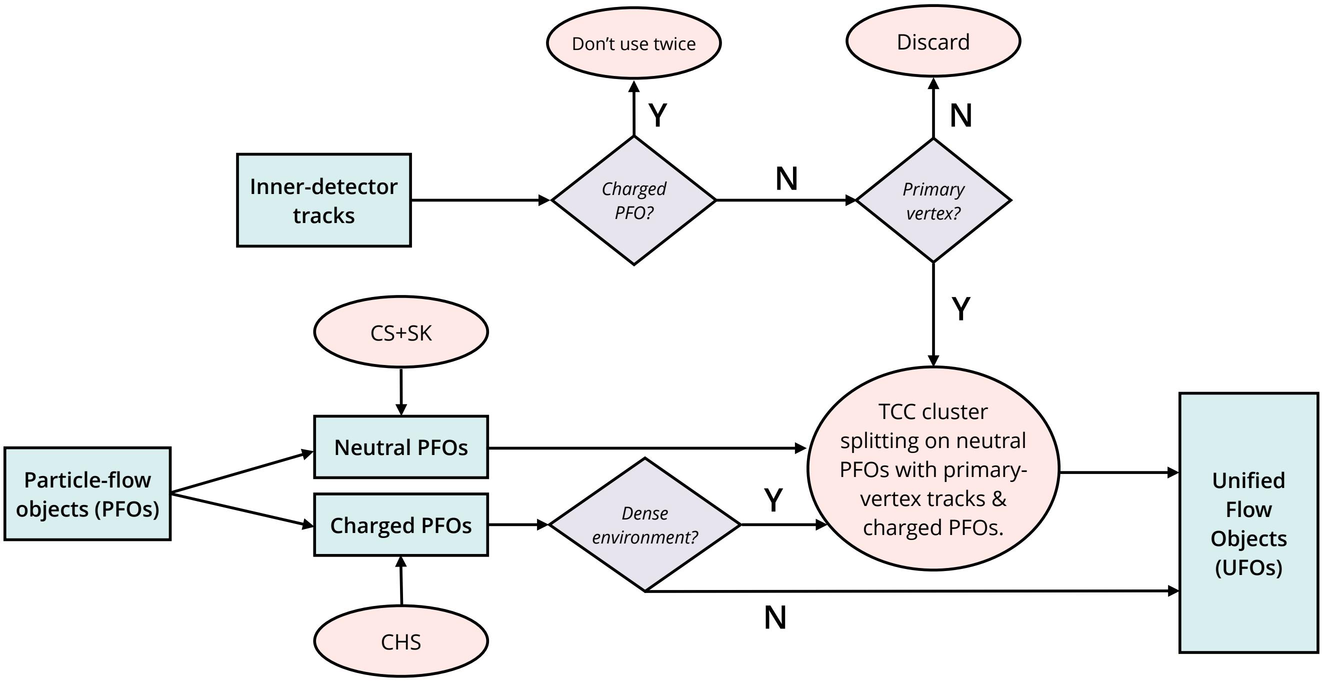


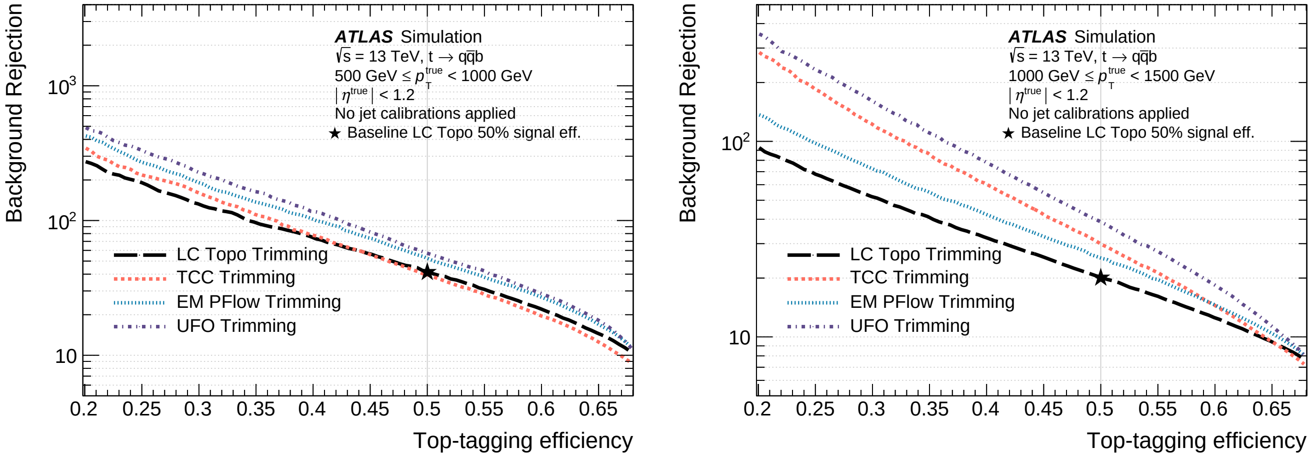
Figure 4.10: Schematic of the Unified Flow Objects algorithm [44].

It starts by applying the PFlow algorithm as described before : charged PFlows that are matched to pile-up vertices are rejected (that corresponds to the Charged Hadron Subtraction (CHS) step that will be discussed in section 4.3.5). A set of PFlows remains that can be catalogued into three categories : neutral PFlows, charged PFlows for which the energy subtraction procedure has been executed and charged PFlows for which no subtraction has been done due to the high p_T value of the track or to its dense environment. Pile-up mitigation algorithms like Constituent-Subtraction (CS) or Soft-Killer (SK) (also discussed in section 4.3.5) can be applied to neutral PFlows.

The TCC algorithm is then applied to the remaining PFlows with a modification : only tracks from the primary vertex will be used during the energy splitting process. Moreover, tracks that correspond to a charged PFlow for which the expected energy has been subtracted from the calorimeter are not considered and remain unchanged. The TCC algorithm is executed as described before with neutral PFlows instead of topo-clusters and with tracks and un-subtracted charged PFlows coming from the primary vertex.

Finally, this process of building jet input constituents allows to benefit from the performance of the PFlow energy subtraction at low p_T and of the cluster energy splitting at high p_T. The performances in terms of tagging or pile-up stabilities for jets that used UFOs as inputs are increased relatively to LCTopo, PFlow or TCC jets across a large range of p_T values.

For example, Figure 4.11 [44] presents the background rejection (i.e the reciprocal of the background-jet tagging efficiency) as a function of signal efficiency for a tagger that uses substructure variables to tag top quark jets, comparing different jet input constituents. It can be seen that the jets built from UFO present superior performances to TCC and PFlow jets across all the p_T range that is considered. Indeed, for UFO jets, the background rejection at a fixed tagging efficiency is always higher than the one corresponding to TCC and PFlow jets.



(a) 500 GeV $< p_T^{\text{true}} <$ 1000 GeV

(b) 1000 GeV $< p_T^{\text{true}} <$ 1500 GeV

Figure 4.11: Background rejection as a function of signal efficiency for a tagger using substructure variables to tag top quark jets. For two different range in top quark p_T^{true}, a comparison between different jet input constituents is shown, including topo-cluster (LC Topo), Track-CaloCluster (TCC), Particle Flow (EM PFlow) and Unified Flow Object (UFO) (the trimming will be described in section 4.4.2) [44].

4.3.5 Pile-up mitigation algorithms

This section describes the different algorithms that can be applied to jet constituents in order to remove pile-up contributions from the signal of interest.

Charged Hadron Subtraction The Charged Hadron Subtraction (CHS) is a technique that can be applied to reduce contribution from in-time pile-up by removing charged constituents which are associated to pile-up vertices and thus will not be used to form jets. In particular, this procedure is applied at the beginning of the UFO algorithm to the charged PFlows that served as inputs.

Constituent Subtraction and Soft-Killer The Constituent Subtraction (CS) algorithm [45] is based on the estimation of the pile-up energy density in the event. This pile-up density ρ is estimated event-by-event as the median of the p_T/A distribution of all jets reconstructed with the k_t algorithm with $R = 0.4$, where A is the jet area in the (η, ϕ) plane (the jet area and the clustering algorithms will be described in section 4.4.1). Then, massless low-p_T virtual constituents are added to the event such as they uniformly cover the (η, ϕ) plane. These constituents are qualified as «ghost constituents», and their transverse momentum p_T^{g} satisfy :

$$p_T^{\mathrm{g}} = A_{\mathrm{g}} \times \rho \tag{4.12}$$

where A_{g} is the area of the ghosts which is set to 0.01. In this way, p_T^{g} accounts for the expected pile-up contribution in a small area $\Delta\eta \times \Delta\phi$ with $\Delta\eta = \Delta\phi = 0.1$. Once the ghosts have been added to the event, the distance $\Delta R_{i,k}$ in the (η, ϕ) plane between each constituents i and ghosts k is calculated. All the constituent-ghost pairs are then sorted in ascending order of $\Delta R_{i,k}$ and the pile-up mitigation algorithm proceeds iteratively by starting from the pair with the lowest $\Delta R_{i,k}$. At each step, the constituent and ghost transverse momentum are modified according to :

$$\text{If } p_{T,i} \geq p_{T,k}^{\mathrm{g}} : p_{T,i} \rightarrow p_{T,i} - p_{T,k}^{\mathrm{g}} \text{ and } p_{T,k}^{\mathrm{g}} \rightarrow 0 \text{ GeV} \tag{4.13}$$

$$\text{Otherwise} : p_{T,i} \rightarrow 0 \text{ GeV and } p_{T,k}^{\mathrm{g}} \rightarrow p_{T,k}^{\mathrm{g}} - p_{T,i} \tag{4.14}$$

until only pairs with $\Delta R_{i,k} > \Delta R_{\max}$ remain, with $\Delta R_{\max}$ set to 0.25 in order that only nearby ghosts are used to correct the transverse momentum of a given constituent [46]. After the procedure, all the ghosts and the constituents with zero transverse momentum are removed from the event.

The Soft-Killer (SK) algorithm [47] consists in applying a cut on the constituent transverse momentum in order to remove pile-up contributions corresponding to low-p_T emissions. The p_T cut value is determined event-by-event. First, the detector is divided into a grid in the (η, ϕ) plane with patches of dimensions equal to $\Delta\eta = \Delta\phi = 0.6$ [46]. Then, the p_T threshold is determined such that after the cut, half of the patches will be empty and will not contain any constituents. With this choice, the resulting pile-up density defined previously satisfies $\rho \sim 0$.

Both CS and SK algorithms are applied in this order at the beginning of the UFO algorithm to the neutral PFlows that served as inputs.

4.4 Jet reconstruction

4.4.1 Jet clustering

The jet clustering algorithms aim to regroup reconstructed constituents to define jets.

In ATLAS, the jet clustering algorithm that is used is the anti-k_t algorithm [48] and it is executed with the FastJet software [49]. It takes as inputs the four-vectors computed for each reconstructed constituents in a given event. It starts by computing the distances :

$$\text{for each pair of constituents : } d_{ij} = \min(1/p_{T,i}^2, 1/p_{T,j}^2) \times (\Delta R_{ij}/R)^2 \tag{4.15}$$

$$\text{for each constituent : } d_{iB} = 1/p_{T,i}^2 \tag{4.16}$$

with $\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ the angular distance between the two constituents i and j and R the radius of the jet which is a free parameter of the algorithm. Then, the algorithm considers the minimum distance among all the d_{ij} and d_{iB} that have been computed. If this minimum is a distance between two constituents i and j, these two constituents are combined to form a new one defined by the addition of the two four-vectors. This entity is then added to the set of remaining constituents while the i and j ones are removed and will not be used during future iterations. If the minimum is one of the d_{iB}, the constituent i (which can come from the combination of several constituents during previous iterations) is considered as a final jet and is removed from the set of constituents. All the possible distances d_{ij} and d_{iB} are then recomputed and the procedure is repeated until no constituents are left.

Given the expression of d_{ij}, this procedure tends to cluster the highest p_T constituents with nearby ones before considering the clustering of pair of low p_T constituents : in this way, the shape of the jets produced by the anti-k_t algorithm is not sensitive to the soft constituents that served as inputs, especially to the ones coming from soft emissions during the parton shower. In the case of a high p_T constituent surrounded only by low p_T ones within a distance of $2R$, the algorithm will simply cluster all the soft constituents within a cone of radius R, producing a jet having a circular shape in the (η, ϕ) plane.

After the clustering procedure, the four-vector of a jet can be computed by summing the four-vectors of the constituents that belong to it, and the jet kinematic variables, such as the energy, the mass, the angular coordinates or the transverse momentum can be deduced. The corresponding constituent four-vectors can also be used to compute jet substructure variables that describe the jet internal energy distribution. Such variables are often used to characterize jets according to the nature of the particle from which they originate (light quark versus gluon, bottom quark or more massive particles such as top quark, W, Z or Higgs boson).

In order to estimate the area A of a jet, which is notably used in the Constituent Subtraction algorithm described in section 4.3.5, a technique named «ghost association» is used. This procedure starts by adding to the considered event virtual constituents named ghosts having an infinitesimal p_T and uniformly covering the (η, ϕ) plane. Then, these ghosts are taking into account during the clustering process in addition to the reconstructed constituents. Because of their infinitesimal momentum, these ghosts do not affect the physics of the jet clustering procedure but are simply combined along physical constituents. Given the fact that each ghost has a known area in the (η, ϕ) plane, the area of a given jet is estimated by summing the area of all the ghosts that have been associated to it. For a jet reconstructed with the anti-k_t algorithm and not overlapping with an other jet, its area verifies $A \approx \pi R^2$, with R the radius parameter.

Two other algorithms, differing only on the definition of d_{ij} and d_{iB}, are also commonly considered to cluster constituents : the k_t algorithm with $d_{ij} = \min(p_{T,i}^2, p_{T,j}^2) \times (\Delta R_{ij}/R)^2$ and $d_{iB} = p_{T,i}^2$ and the Cambridge/Aachen algorithm with $d_{ij} = (\Delta R_{ij}/R)^2$ and $d_{iB} = 1$. Regarding these definitions, the k_t algorithm tends to cluster first pair of soft constituents while the Cambridge/Aachen one regroup in priority closest constituents no matter what are their transverse momentum.

Moreover, these jet clustering algorithms can be applied in the same way to truth particles obtained from MC simulations as inputs instead of reconstructed constituents. These truth particles correspond to the stable final-state ones (except muons, neutrinos and particles from pile-up interactions) obtained after the simulation of the hadronization, and before the simulation of the interaction with the detector. In this case, the jets obtained are named «truth jets» while the ones built from reconstructed constituents are named «reconstructed jets». In MC based studies, the quantities calculated from truth jets such as the energy or the mass serve as reference for the corresponding reconstructed quantities.

In ATLAS, the most common values for the radius R are either $R = 0.4$ or $R = 1.0$. The former is used to reconstruct jets initiating by light quarks or gluons (small-R jets), and the latter is used to reconstruct boosted hadronically decaying W, Z, Higgs bosons or top quarks (large-R jets).

4.4.2 Jet grooming algorithms

Using large-R jets allow to efficiently capture the majority of the decay products. However, these jets also capture more particles from pile-up or from the underlying event, strongly impacting the calculation of jet substructure variables. To mitigate these sources of contamination, grooming algorithms can be applied to the reconstructed large-R jets. Moreover, these algorithms also allow to reconstruct better the jet energy and mass with respect to the truth jet ones.

The first algorithm used in ATLAS is named «Trimming» [50]. The procedure is divided into several steps. First, for each reconstructed jet, the corresponding constituents are reclustered into sub-jets with a radius R_{sub} smaller than the radius of the jet (the algorithm used for the reclustering can be different from the one used during the clustering into large-R jets). Secondly, the algorithm discards the sub-jets having a transverse momentum such as $p_T^{\text{sub}} < f_{\text{cut}} \times p_T$, with f_{cut} corresponding to a parameter of the algorithm and p_T being the transverse momentum of the jet that is considered. Finally, the constituents from the remaining sub-jets form the trimmed jet. The parameter values commonly used in ATLAS are $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.05$, and it is the k_t algorithm that is generally used for the reclustering.

The second algorithm which is generally considered is named «Soft-drop» [51]. Such as in the trimming algorithm, the Soft-drop procedure starts by the reclustering of the constituents of a given jet. It is done with the Cambridge/Aachen algorithm in order to firstly regroup the closest constituents together. Secondly, the jet j is divided into two sub-jets, j_1 and j_2, corresponding to the ones that form the last pair during the Cambridge/Aachen clustering. If the two sub-jets satisfy the condition : $\min(p_{T,1}, p_{T,2})/(p_{T,1} + p_{T,2}) > z_{\text{cut}}(\Delta R_{12}/R)^{\beta}$, with $p_{T,i}$ the transverse momentum of the sub-jet i, ΔR_{12} the distance between the two sub-jets in the (η, ϕ) plane, R the radius of the jet and z_{cut} and β two parameters of the algorithm, then the jet j is the final soft-dropped jet. Otherwise, the sub-jet with the lowest p_T is discarded and j is redefined as the sub-jet with the highest p_T, and the condition is verified again on this new jet. The procedure is repeated until the inequality described previously is satisfied : the remaining constituents form the soft-dropped jet. The parameter values commonly used in ATLAS are $z_{\text{cut}} = 0.1$ and $\beta = 1$.

4.4.3 Jet calibration

Due to detector effects like the energy losses in non active materials or the non-compensating nature of the calorimeters, the reconstructed jet kinematic observables are in general different from the truth jet ones. For example, the jet energy response, defined as the ratio between the reconstructed jet energy and the truth jet one, is generally different from one. As a consequence, the reconstructed jets are corrected and calibrated in order to coincide with truth jets. This jet calibration is done in several steps that are schematized in Figure 4.12 and that will briefly summarized below [52].

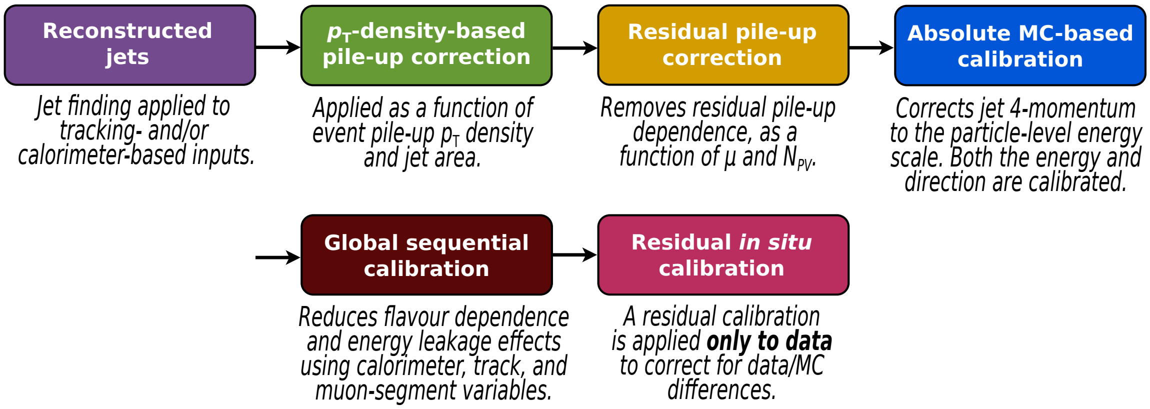


Figure 4.12: Schematic of the jet calibration [52].

With the exception of the «Residual in situ calibration» which is derived from data, all the steps of the calibration chain are derived from MC simulations. Indeed, to derive the different corrections that must be applied, it is necessary to have a truth correspondence for each reconstructed jets. It is done by simulating the interactions with the detector of the truth particles thanks to Geant4 [34] : the reconstructed jets from MC simulations can then be formed in the same way as the ones from data signals, and each of them can be matched to a truth jet.

The first step of the calibration is a correction that aims to remove the energy contribution coming either from in-time or out-of-time pile-up. Two different corrections are applied to the transverse momentum of the reconstructed jet : the first one is applied as a function of the jet area and the pile-up density (computed in a similar way as in the CS algorithm), while the second one is evaluated as a function of the number of primary vertices in the event N_{PV} and the mean number of interactions per bunch crossing μ, which are variables respectively quantifying the in-time and the out-of-time pile-up. These two corrections correspond to the «p_T-density-based pile-up correction» and «Residual pile-up correction» blocks in Figure 4.12.

Secondly, the reconstructed jet four-vector is calibrated with a correction of the jet energy and η position. The calibration factors have been derived from simulations as a function of the reconstructed jet energy and η. More precisely, each reconstructed jet obtained from simulation is matched to a truth jet, before distributions of the energy response $E_{\text{reco}}/E_{\text{true}}$ or the η response $\eta_{\text{reco}} - \eta_{\text{true}}$ being derived in bins of E and η. From these distributions, the jet energy scale factors and η corrections are evaluated. This step corresponds to the «Absolute MC-based calibration» block in Figure 4.12.

Because differences between the reconstructed and truth jets can also be due to the flavor and energy distribution of the constituents for instance, an additional calibration step corresponding to a series of multiplicative corrections is executed to reduce such dependencies, while not changing in average the corrected jet energy obtained from the «Absolute MC-based calibration». Theses corrections are derived as a function of several jet observables that either describe the longitudinal structure of the energy deposited in the calorimeters, the tracks associated to the jet or the activity in the muon chambers behind the jet. This step is named «Global sequential calibration» in Figure 4.12.

Finally, the last step of the calibration is applied only to jets reconstructed from data, and aims to correct the remaining differences between data and MC simulation that can come from the imperfect simulation of the particle interactions with the detector. This correction is determined by exploiting the transverse momentum conservation in events containing one jet and a well-reconstructed object, such as a photon or a leptonically decaying Z boson. Momentum imbalance between the jet and the reference object can be evaluated in simulation and in data, and the differences are interpreted as corrections that need to be applied to data jets. This step corresponds to the «Residual in situ calibration» block in Figure 4.12.

All of these steps are applied on small-R jets but it is not the case for large-R jets : concerning the latter, the «Global sequential calibration» and the two pile-up corrections are not applied since grooming algorithms are executed.

4.5 Improvement on UFO jet performances

Studies on small-R jet performances have demonstrated small degradations for small-R jets built from UFO constituents with respect to PFlow jets, especially for jets located at the end of the inner detector acceptance ($|\eta| = 2.5$). For example, the Figure 4.13 [53] shows the jet energy resolution (defined in section 4.5.3) after the jet energy scale calibration as a function of the truth jet $|\eta|$ for small-R jets built either from UFO or PFlow constituents. It can be seen that the jet energy resolution is similar between the two types of jets except between $|\eta| = 2.4$ and $|\eta| = 2.8$ where the jet energy resolution is poorer for UFO jets.

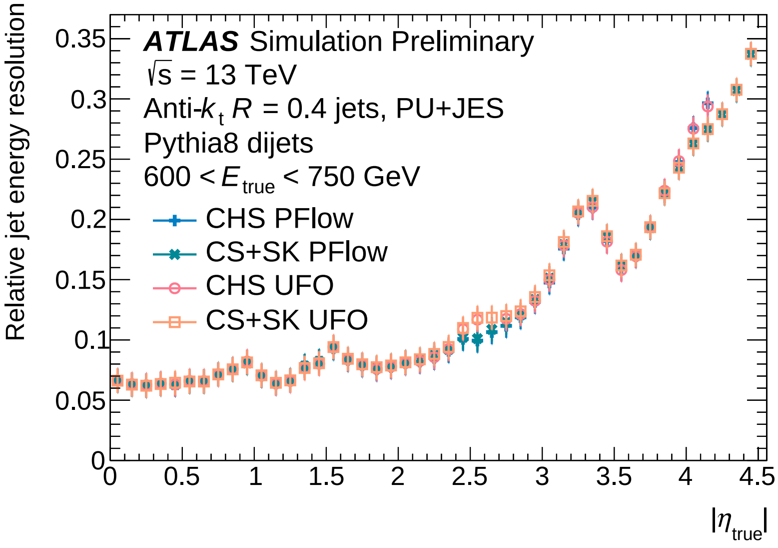


Figure 4.13: Relative jet energy resolution after the jet energy scale calibration as a function of the truth jet $|\eta|$ for small-R jets built from UFO and PFlow constituents [53].

The ATLAS authorship qualification task that I worked on was to understand the origin of this unexpected degradation and to explore different ways of improving the UFO reconstruction algorithm.

4.5.1 MC samples and event reconstruction

The UFO jet performances are studied by comparing UFO jets reconstructed from simulations to the corresponding truth jets. The MC samples used for the studies have been produced using Pythia 8.235 [30] with the NNPDF2.3LO parton distribution functions set [54] and the A14 set of tuned parameters [55] at a center-of-mass energy of $\sqrt{s} = 13$ TeV.

The physics process that has been simulated corresponds to the production of two partons in the final state, leading to QCD di-jet events. These di-jet events have been generated in slices of truth jet p_T in order to have a sufficient statistic across a large kinematic range (from 60 to 4600 GeV).

The jet input constituents that are used are :

- the UFO on which the CHS and CS+SK algorithms are applied,
- the PFlow on which only the CHS procedure is applied.

They are clustered into jets with the anti-k_t algorithm thanks to the FastJet software with an angular radius of 0.4. For each event, in addition to the truth jets, two collections of reconstructed small-R jets are thus considered : UFO and PFlow jets.

4.5.2 Selections

The events are selected to have at least two reconstructed jets. To be considered in the following studies, each reconstructed jet (UFO or PFlow) must be matched to a truth jet. A reconstructed jet is matched to a truth one if the angular distance that separates them satisfies $\Delta R < 0.3$. Moreover, each reconstructed jet must not have any other reconstructed jet within a distance of $\Delta R < 0.6$. In addition, a reconstructed jet is not considered if it is matched to a truth jet having an other truth jet within a distance of $\Delta R < 1.0$.

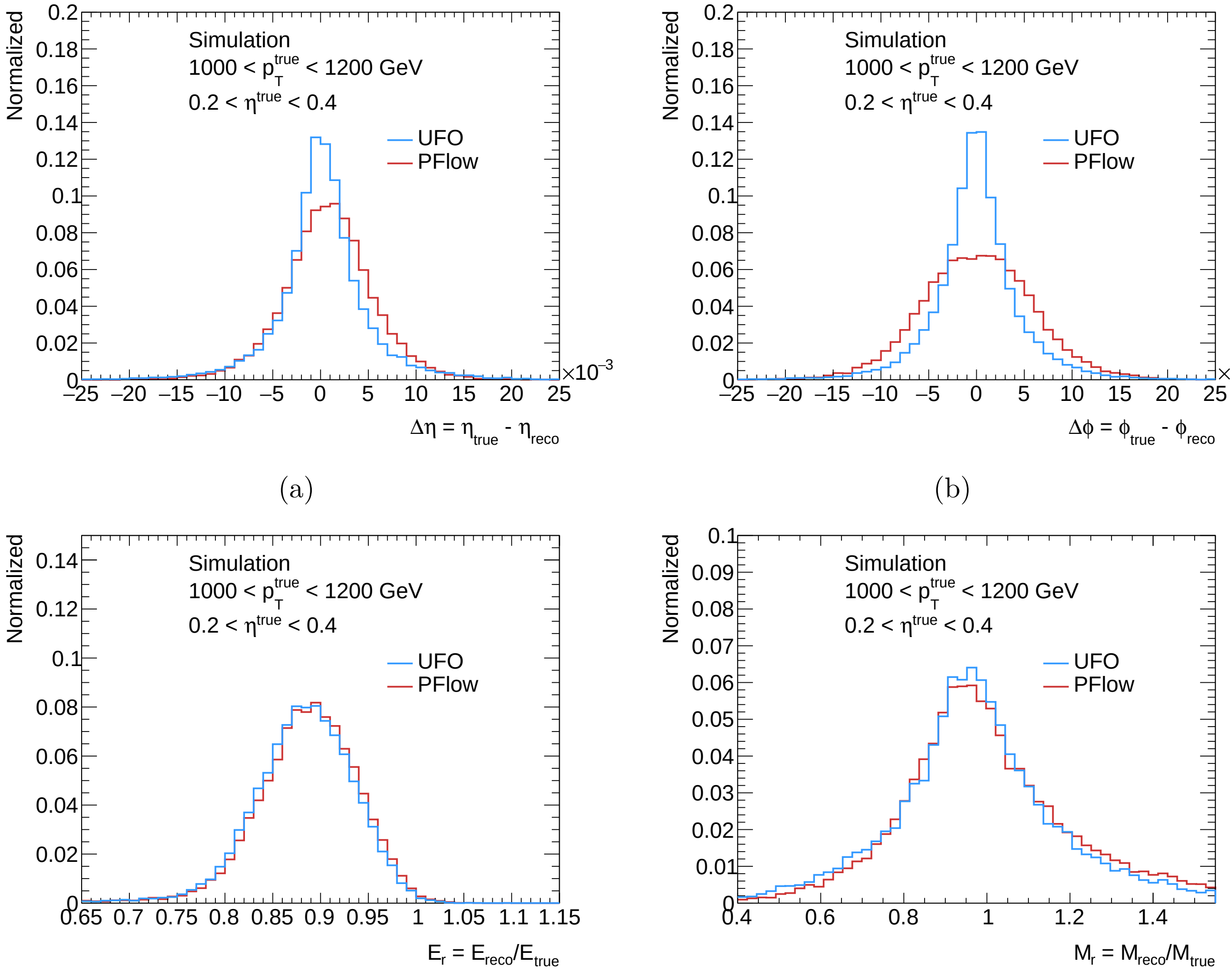
These isolation requirements allow to ensure the samples studied consist essentially of QCD jets from high-momentum partons, which are not contaminated by hadronic activity from pile-up or the underlying event. Indeed, the modelling of the latter is poorer, and such activity could thus lead to bias in the studies.

4.5.3 UFO and PFlow jet performances

The UFO and PFlow jet performances are evaluated by considering the jet energy, mass and angular resolutions. To evaluate these resolutions, distributions of the variables $\Delta\eta = \eta_{\text{true}} - \eta_{\text{reco}}$, $\Delta\phi = \phi_{\text{true}} - \phi_{\text{reco}}$, $E_r = E_{\text{reco}}/E_{\text{true}}$ and $M_r = M_{\text{reco}}/M_{\text{true}}$ are considered : η and ϕ correspond to the jet angular coordinates, E and M to the jet energy and mass, «reco» designates the reconstructed jet before any calibration while «true» accounts for the truth jet. These variables are respectively named the *individual jet η, ϕ, energy and mass responses*, and their distributions are derived in bins of truth jet p_T and η. In Figure 4.14, normalized distributions of $\Delta\eta$, $\Delta\phi$, E_r and M_r are represented for both UFO and PFlow jets for one interval in η and p_T.

Then, the jet η, ϕ, energy and mass resolutions are defined as the inter-quartile range at 68 % of their respective response distributions. They quantify the width of these distributions around the most probable value, and thus are desired to be as small as possible.

Performances at the end of the tracker acceptance The performances in the forward region of the detector ($\eta \geq 2.5$) can only be considered with low-p_T (below $\sim$ 600 GeV) small-R jets, because high-p_T jets tend to have central η values (cf Figure 4.15).



(a) (b)

(c) (d)

Figure 4.14: Normalized distributions of (a) $\Delta\eta$, (b) $\Delta\phi$, (c) E_r and (d) M_r for UFO and PFlow jets.

Figure 4.16 shows the distributions of $\Delta\eta$ for UFO and PFlow jets in bins with η around ± 2.5. The degradation for UFO jets is highlighted in these distributions which are larger for UFO jets compared to the PFlow jet ones.

Moreover, in bins with η close to $2.5(-2.5)$, the UFO distributions are displaced to positive (negative) values while the distributions for PFlow jets are rather centered at 0. Because $\Delta\eta = \eta_{\text{true}} - \eta_{\text{reco}}$, it demonstrates that there is an amount of UFO jets located around $|\eta| = 2.5$ for which the reconstructed $|\eta|$ is significantly inferior to the truth value, while this behavior is not observed for PFlow jets.

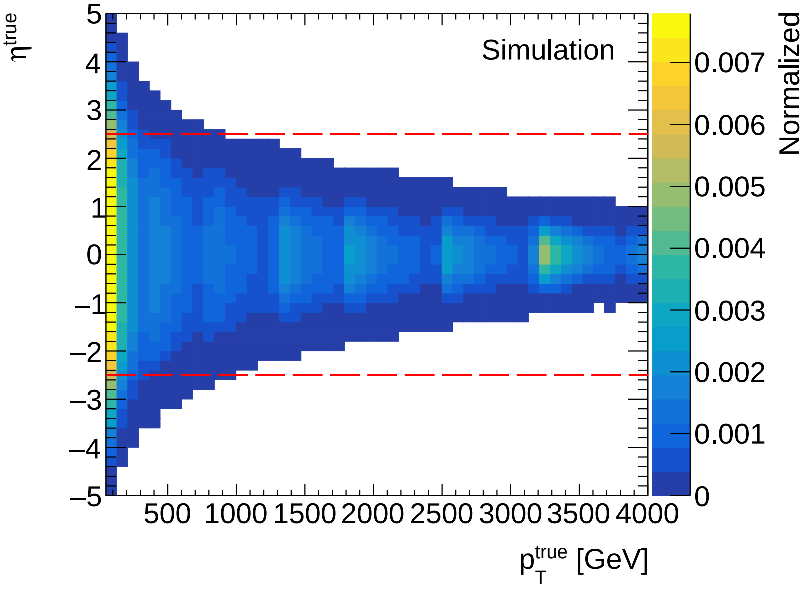


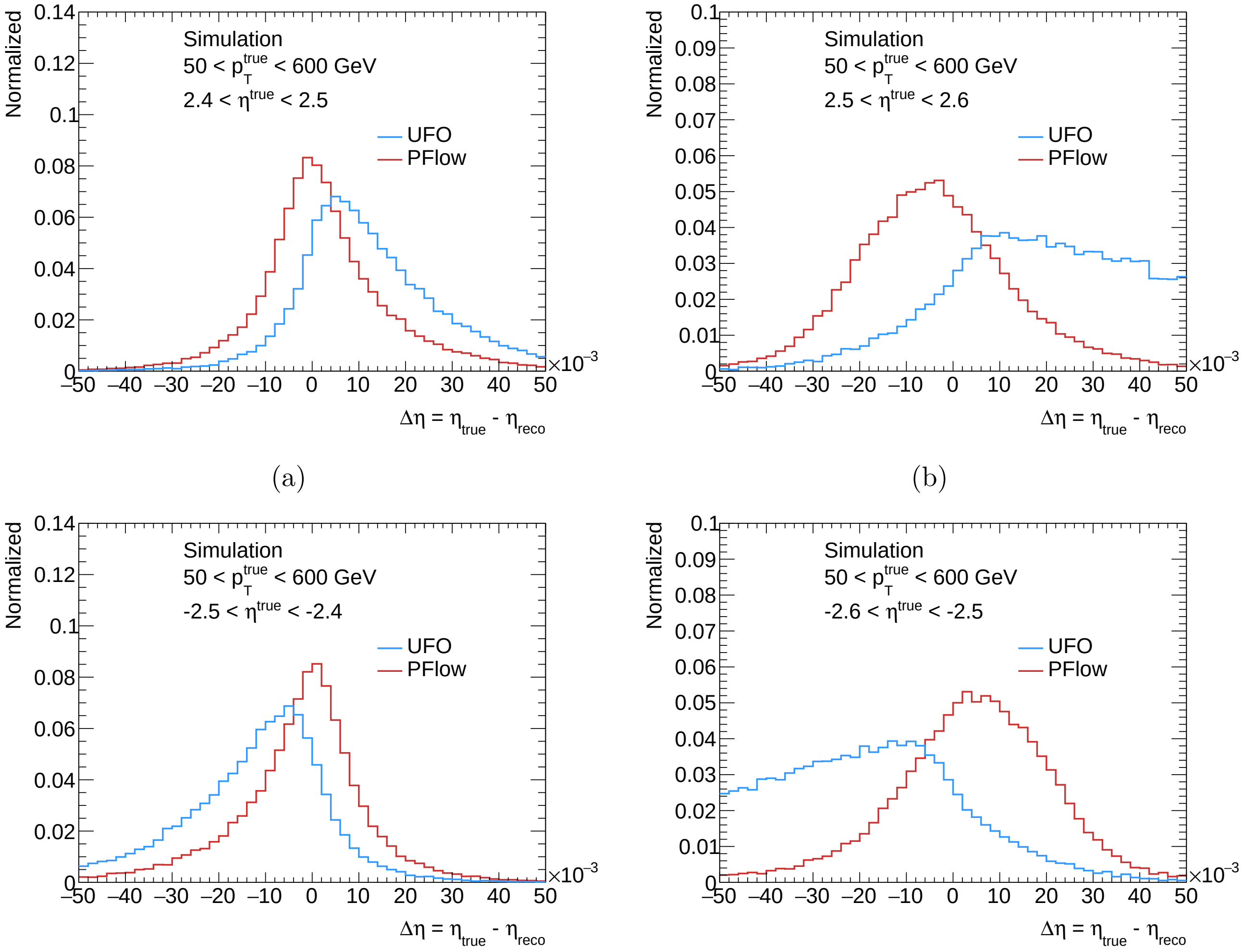
Figure 4.15: Normalized 2D distributions of the small-R jets in the plane $(p_T^{\text{true}}, \eta^{\text{true}})$. The red dashed lines at $|\eta| = 2.5$ indicate the end of the inner detector coverage.

The main difference between jets built from PFlow and UFO constituents is the possibility for the latter to contain combined constituents, which are constructed by combining information from a track and one or more clusters. Consequently, the possible correlation between the jet angular response and the *jet energy fraction of combined constituents*, given by $\sum e_{\text{constituent}}/E_{\text{jet}}$ with the sum occurring on the combined constituents of the jet, has been explored.

In Figure 4.17, the jet η resolution for UFO jets is shown as a function of the UFO jet energy fraction of combined constituents in (a) the central part of the detector and (b) the forward region corresponding to $|\eta| > 2.5$. The resolution for PFlow jets can also be shown as a function of the corresponding UFO jet energy fraction of combined constituents, as each PFlow jet and its corresponding UFO jet are matched to the same truth jet.

In the forward region, it can be observed that that disparities between UFO and PFlow jets appear with increasing combined energy fraction : for UFO jets that are mainly composed of combined constituents, the jet angular response is significantly worse compared to PFlow jets.

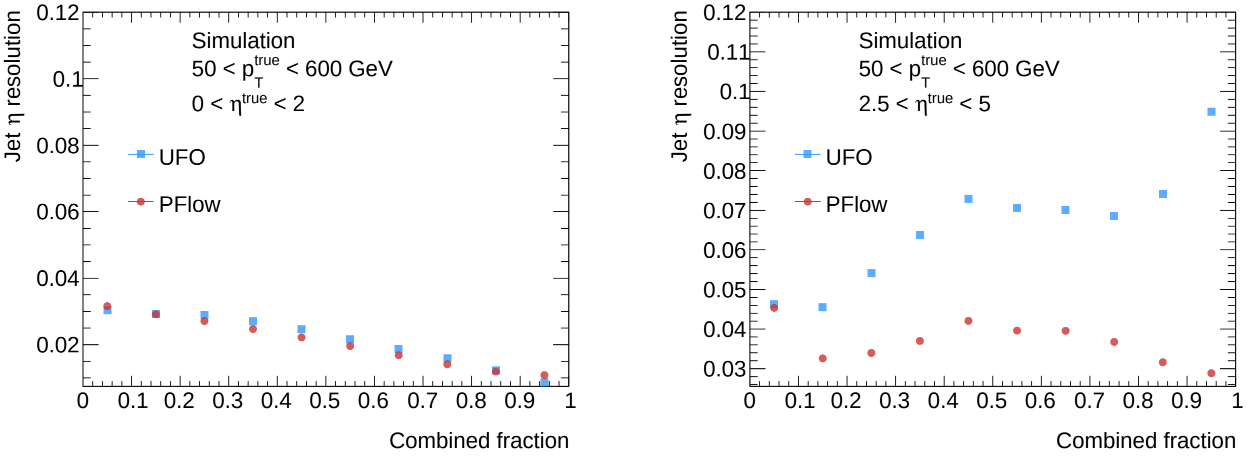
On the other hand, in the central region of the detector, the jet η resolution is similar between PFlow and UFO jets, even with UFO jets having an important energy fraction of combined constituents. It thus confirms that the observed degradations in the forward region are related to the UFO combined constituents.



(a) (b)

(c) (d)

Figure 4.16: Normalized distributions of $\Delta\eta$ for UFO and PFlow jets in bins corresponding to η^{true} values close to the end of the inner detector coverage.

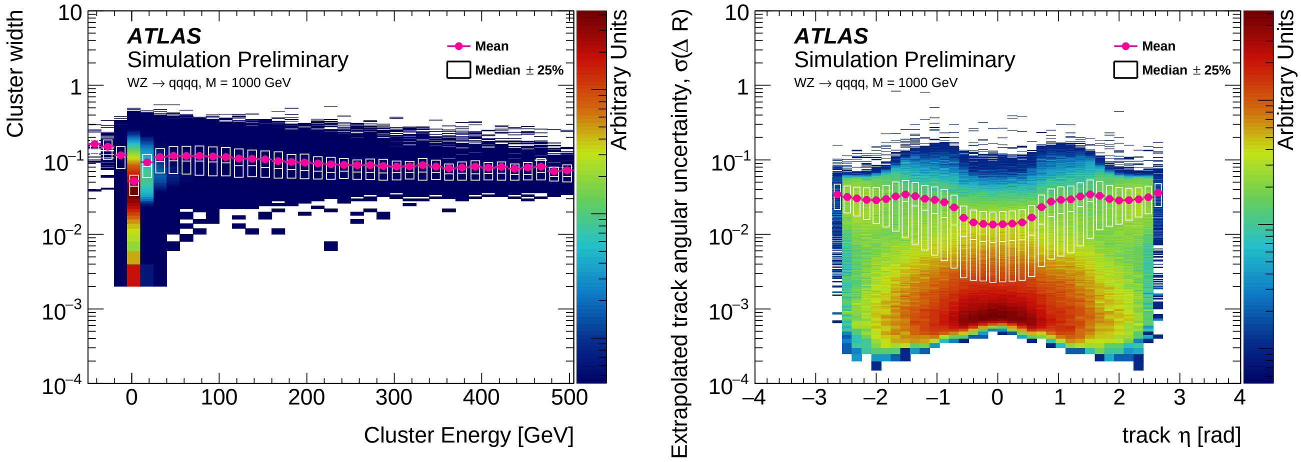


(a) (b)

Figure 4.17: Jet η resolution as a function of the UFO energy fraction of combined constituents for jets located (a) in the central part of the detector and (b) in the forward region. For each PFlow jet, the combined fraction that is considered is the one from the corresponding UFO jet, i.e the one that is matched to the same truth jet.

4.5.4 Modified TCC algorithm

As described in section 4.3.3, each combined constituent always inherits its angular coordinates η and ϕ from the corresponding track. This allows to explain the fact that a UFO jet located around the end of the inner detector acceptance and that contains a significant amount of combined constituents can have a global η value that is brought to the central part of the detector.



(a)

(b)

Figure 4.18: (a) Topo-cluster width as a function of the topo-cluster energy and (b) uncertainty on the extrapolation of a track to the calorimeter as a function of the track η coordinate, all evaluated in $W' \to WZ \to \bar{q}q\bar{q}q$ events with $m_{W'} = 1$ TeV [43].

Indeed, clusters located at $|\eta| \geq 2.5$ can only be matched to tracks with an $|\eta|$ coordinate strictly below 2.5. However, a track is matched to a cluster if the angular distance in (η, ϕ) between the track and the cluster satisfies $\Delta R < \Delta R_{\text{cut}} = \sqrt{\sigma^2_{\text{cluster}} + \sigma^2_{\text{track}}}$, with σ_{cluster} the topo-cluster width and σ_{track} the track extrapolation uncertainty. In this way, in the case of a high energy cluster having a width about 0.1 and its barycentre located at $\eta \sim 2.55$, a track located at $\eta \sim 2.45$ can be matched to it, changing the combined η position regarding the topo-cluster one, and thus notably affecting and decreasing the η position of the UFO jet that will contain this constituent.

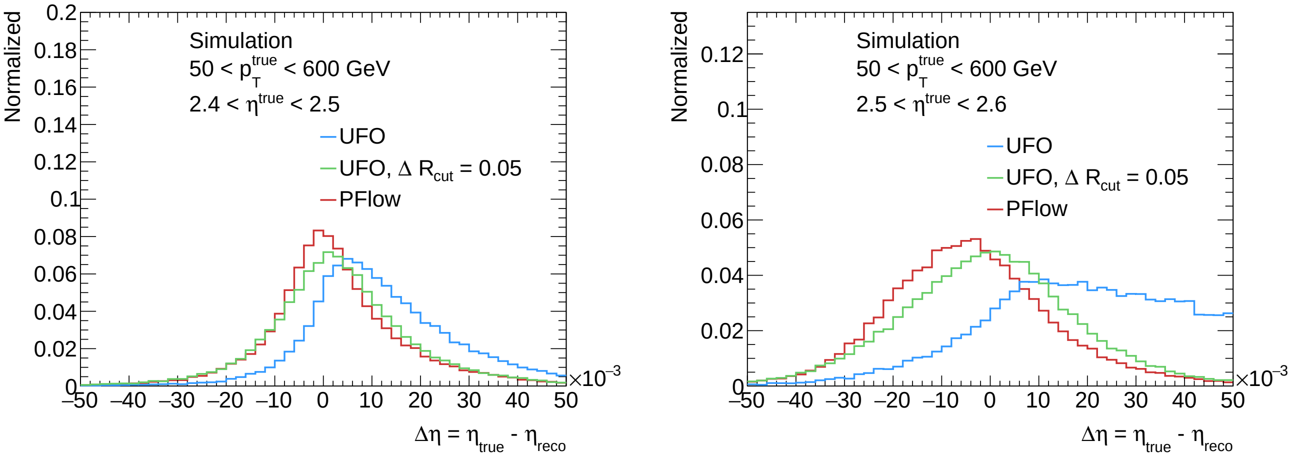
In order to possibly improve the jet angular reconstruction performances, a fixed value for ΔR_{cut} independent from σ_{cluster} or σ_{track} has been tested. Considering the dependence of the cluster width and track extrapolation uncertainty as a function of the cluster energy and track η coordinate (cf Figure 4.18.a and 4.18.b), the tested cut value has been set to $\Delta R_{\text{cut}} = 0.05$. The motivation for this choice is that this value remains higher than the track extrapolation uncertainty in average, while being more restrictive for clusters having an energy higher than 20 GeV, for which the mean cluster width is around 0.1.

To evaluate the impact of this change in the TCC algorithm, modified UFO constituents are built based on this new track-cluster matching condition before being clustered again into small-R jets.

4.5.5 Performances of modified UFO jets

Distributions of $\Delta\eta$ for jets around $\eta = 2.5$ are shown in Figure 4.19, including the standard UFO jets, the modified UFO jets and the PFlow jets. It is directly possible to see the benefit of this new track-cluster matching condition since modified UFO jets present improved performances with respect to standard ones and behave similarly as PFlow jets.

The jet η and ϕ resolutions are then shown as a function of η^{true} in Figure 4.20.a and 4.20.b. These figures also show that the modification in the TCC algorithm allows to obtain similar resolutions between UFO and PFlow jets especially around $\eta = 2.5$.

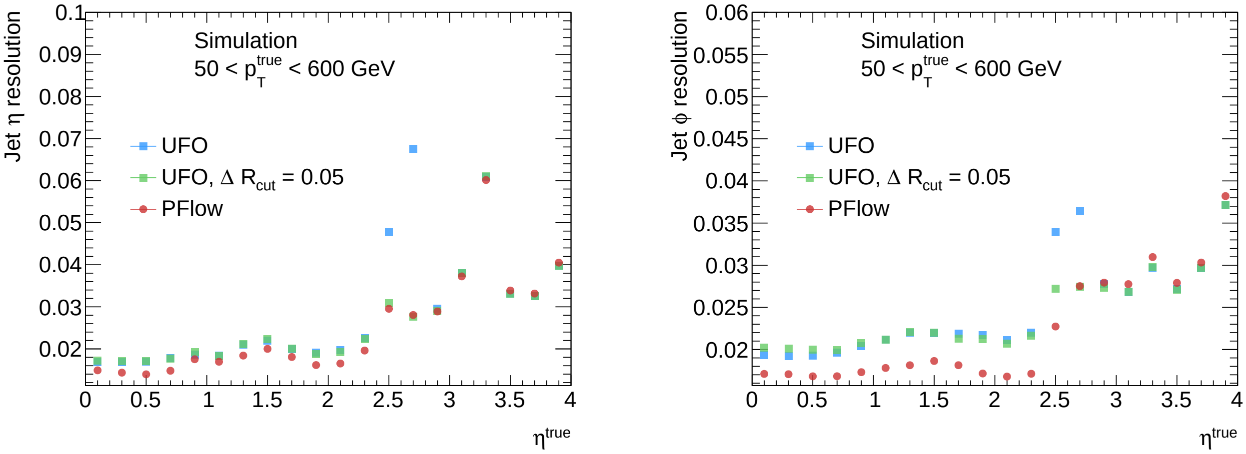


(a)

(b)

Figure 4.19: Normalized distributions of $\Delta\eta$ for UFO (both standard (blue) and modified (green)) and PFlow jets (red) in bins corresponding to η^{true} values close to the end of the inner detector coverage.

The jet energy (mass) resolution is shown as a function of p_T^{true} in Figure 4.21 (4.22) in (a) the central region of the detector and (b) the forward region across a more restrictive p_T interval. It can be noticed that in the η range that is covered by the inner detector, the modified UFO jet performances are at least as good as the standard UFO ones, while in the forward region, the jet energy and mass resolutions of the modified UFO jets are improved with respect to standard ones.

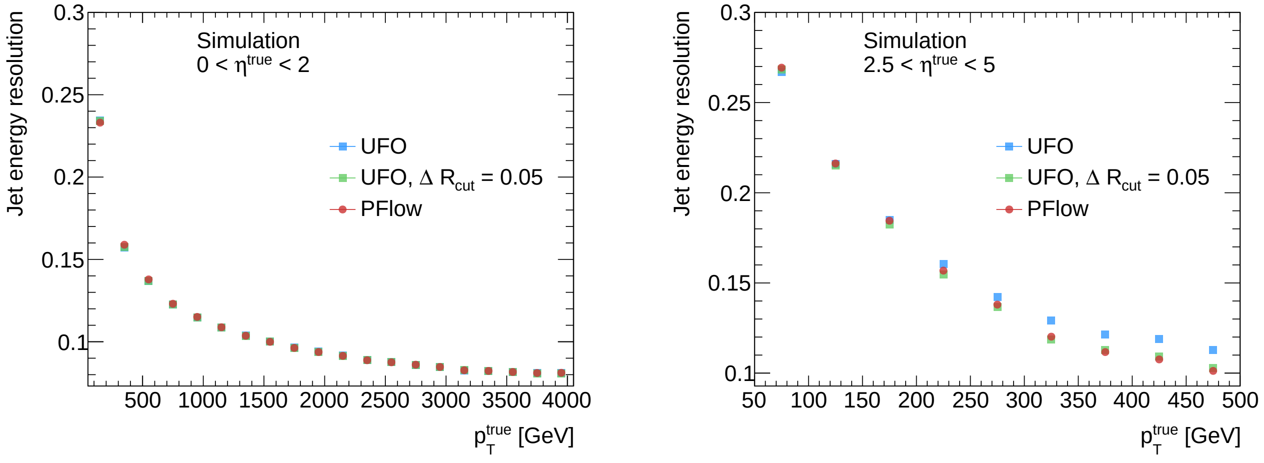


(a) (b)

Figure 4.20: Jet (a) η and (b) ϕ resolutions as a function of η^{true} for UFO (both standard and modified) and PFlow jets.

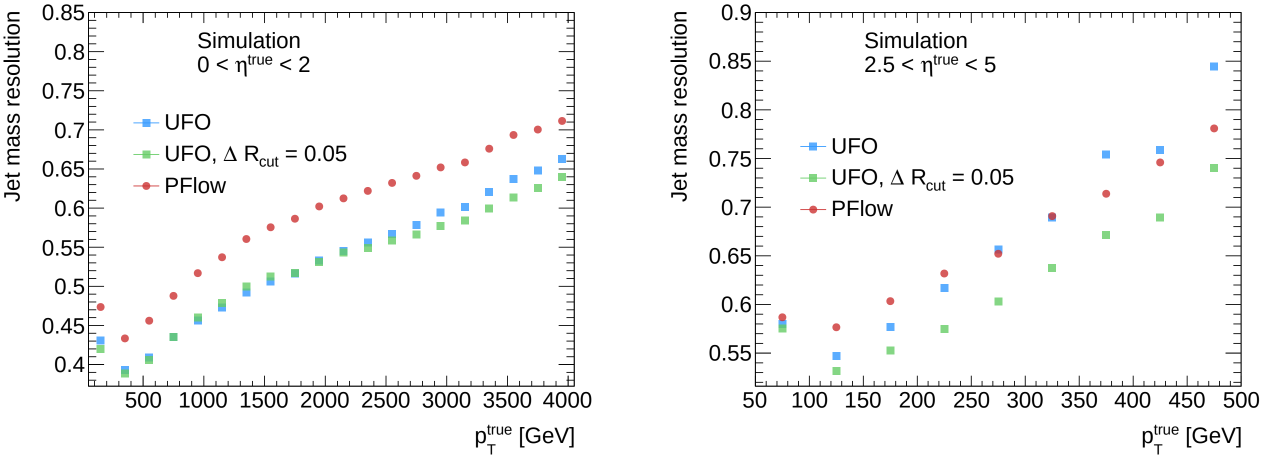
The modification to the TCC algorithm induces a fraction of combined constituents to convert into neutral constituents as shown in Figure 4.23.a and 4.23.b, which represent respectively the mean jet energy fraction of combined constituents and the mean jet energy fraction of neutral constituents as a function of η^{true} for both standard and modified UFO jets. For these constituents, the four-vector is computed differently (cf section 4.3.3), allowing to improve the global jet angular, energy and mass resolutions, especially at the exterior of the inner detector acceptance.

Moreover, this modification does not induce any loss of energy in the global jet reconstruction. Indeed, the mean jet energy responses (i.e the mean values of the E_r distributions) are similar between the standard and modified UFO jets as shown in Figure 4.24.



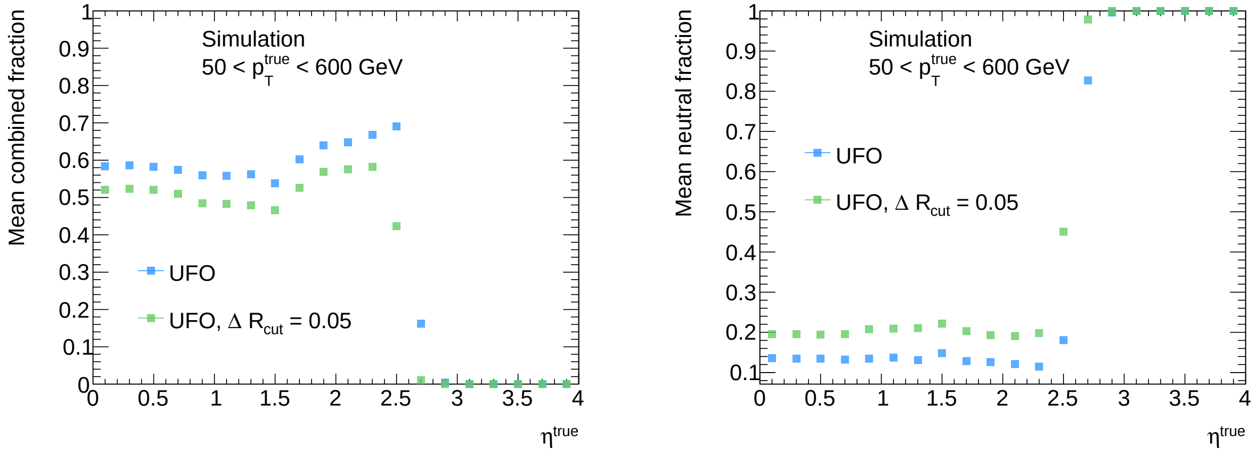
(a) (b)

Figure 4.21: Jet energy resolution as a function of p_T^{true} for UFO (both standard and modified) and PFlow jets located (a) in the central part of the detector and (b) in the forward region.



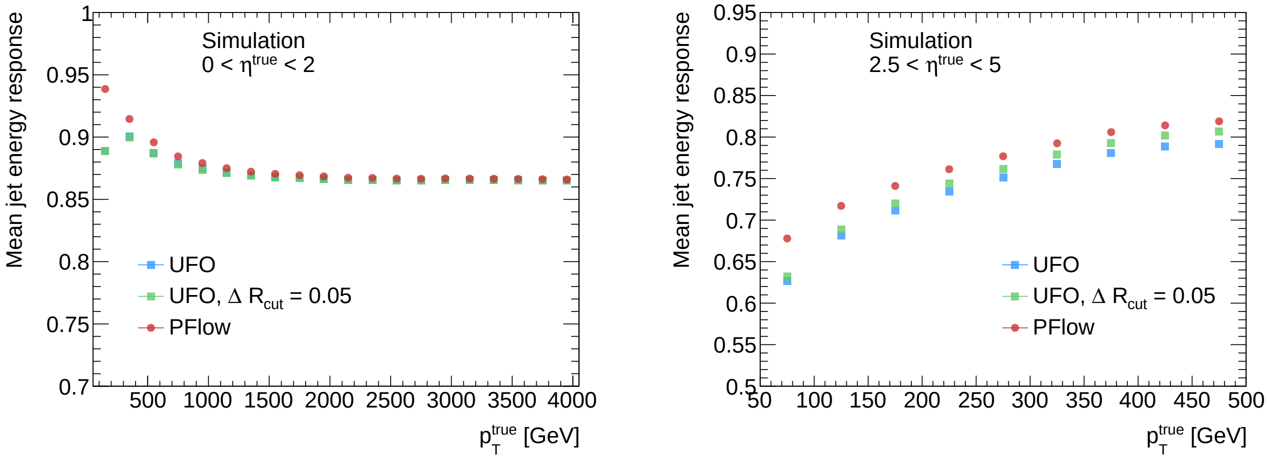
(a) (b)

Figure 4.22: Jet mass resolution as a function of p_T^{true} for UFO (both standard and modified) and PFlow jets located (a) in the central part of the detector and (b) in the forward region.



(a) (b)

Figure 4.23: Mean jet energy fraction of (a) combined and (b) neutral constituents for UFO jets (both standard and modified) as a function of η^{true}.

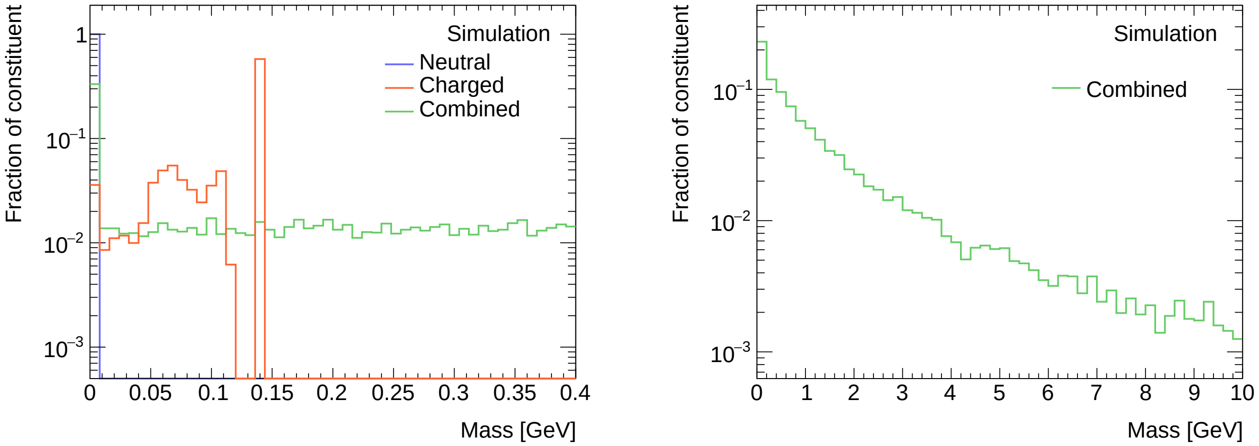


(a) (b)

Figure 4.24: Mean jet energy response as a function of p_T^{true} for UFO (both standard and modified) and PFlow jets located (a) in the central part of the detector and (b) in the forward region.

4.5.6 UFO constituent mass

As explained in section 4.3.3, the TCC algorithm allows UFO combined four-vectors to have a non-zero mass when a track is geometrically matched to more than one cluster. In Figure 4.25, the distribution of the mass for the three different types of UFO constituent is shown. It is possible to see that combined constituents can have large mass, i.e about 10 GeV, which seems physically not consistent and may lead to mis-reconstruction of the global UFO jet mass.



(a) (b)

Figure 4.25: Normalized distributions of the UFO constituent mass for (a) charged, neutral and combined and (b) only combined constituents.

Consequently, an additional test has been performed to possibly improve the UFO jet mass reconstruction. It consists in manually setting the UFO constituent mass to 0. For a given constituent, it is done by computing its momentum components p_x, p_y and p_z based on its energy e and its angular position η and ϕ, by assuming that $|\vec{p}|^2 = p_x^2 + p_y^2 + p_z^2 = e^2$, and using the equations $p_y/p_x = \tan(\phi)$ and $p_z/|\vec{p}| = \tanh(\eta)$. In this way, a massless four-vector with the same energy and angular position can be obtained for each UFO constituent.

Then, the massless constituents are clustered into jets. Unlike the previous studies, it is large-R jets ($R = 1.0$) that are reconstructed to evaluate the impact of this modification in the reconstruction of the UFO. The event selection is similar, but the isolation requirements are different because they rely on the jet radius. In this case, each reconstructed jet must not have any other reconstructed jet within a distance of $\Delta R < 1.5$, and not be matched to a truth jet having an other truth jet within a distance of $\Delta R < 2.5$.

The impact of this procedure has been evaluated by considering distributions of the jet mass and jet substructure variables for standard UFO jets, UFO jets built from massless constituents and truth jets. Several substructure variables have been considered such as Q_w, τ_{21}, τ_{32} and D_2, which are able to discriminate QCD jets from boosted jets coming from top quarks, W or Higgs bosons.

To compute Q_w, the jet is decomposed into three sub-jets using the k_t algorithm. Then, Q_w corresponds to : $\min(m_{12}, m_{13}, m_{23})$, where m_{ij} is the dijet invariant mass computed from the two sub-jet i and j. To evaluate τ_{21} and τ_{32}, the jet N-subjettiness τ_N with $N = 0,\ 1,\ 2,\ 3$, that describes to what degree the substructure of a given jet is compatible with being composed of N or fewer sub-jets, are evaluated [56]. To evaluate τ_N, the jet is first decomposed into N sub-jets using the k_t algorithm. Then, the 0-, 1-, 2- and 3-subjettiness are computed as :

$$\tau_0 = \sum_{i \in J} p_{T_i} R, \tag{4.17}$$

$$\tau_1 = \frac{1}{\tau_0} \sum_{i \in J} p_{T_i} \Delta R_{1,i}, \tag{4.18}$$

$$\tau_2 = \frac{1}{\tau_0} \sum_{i \in J} p_{T_i} \min(\Delta R_{1,i}, \Delta R_{2,i}), \tag{4.19}$$

$$\tau_3 = \frac{1}{\tau_0} \sum_{i \in J} p_{T_i} \min(\Delta R_{1,i}, \Delta R_{2,i}, \Delta R_{3,i}), \tag{4.20}$$

with the sum occurring on the constituents of the jet, R the radius of the jet and $\Delta R_{j,i}$ the angular distance between the sub-jet j and the constituent i. Finally, τ_{21} corresponds to τ_2/τ_1 and τ_{32} to τ_3/τ_2. These variables have demonstrated a great ability to identify respectively two-body and three-body structures within jets. To compute D_2, the N-point energy correlation functions ECFN with $N = 1,\ 2,\ 3$ are first evaluated as :

$$\text{ECF1} = \sum_{i \in J} p_{T_i}, \tag{4.21}$$

$$\text{ECF2} = \sum_{i<j \in J} p_{T_i} p_{T_j} \Delta R_{ij} \tag{4.22}$$

$$\text{ECF3} = \sum_{i<j<k \in J} p_{T_i} p_{T_j} p_{T_k} \Delta R_{ij} \Delta R_{ik} \Delta R_{jk}, \tag{4.23}$$

with the sum occurring on the constituents of the jet and $\Delta R_{i,j}$ the angular distance between the two constituents i and j [56]. Then, dimensionless ratios of these quantities are defined as :

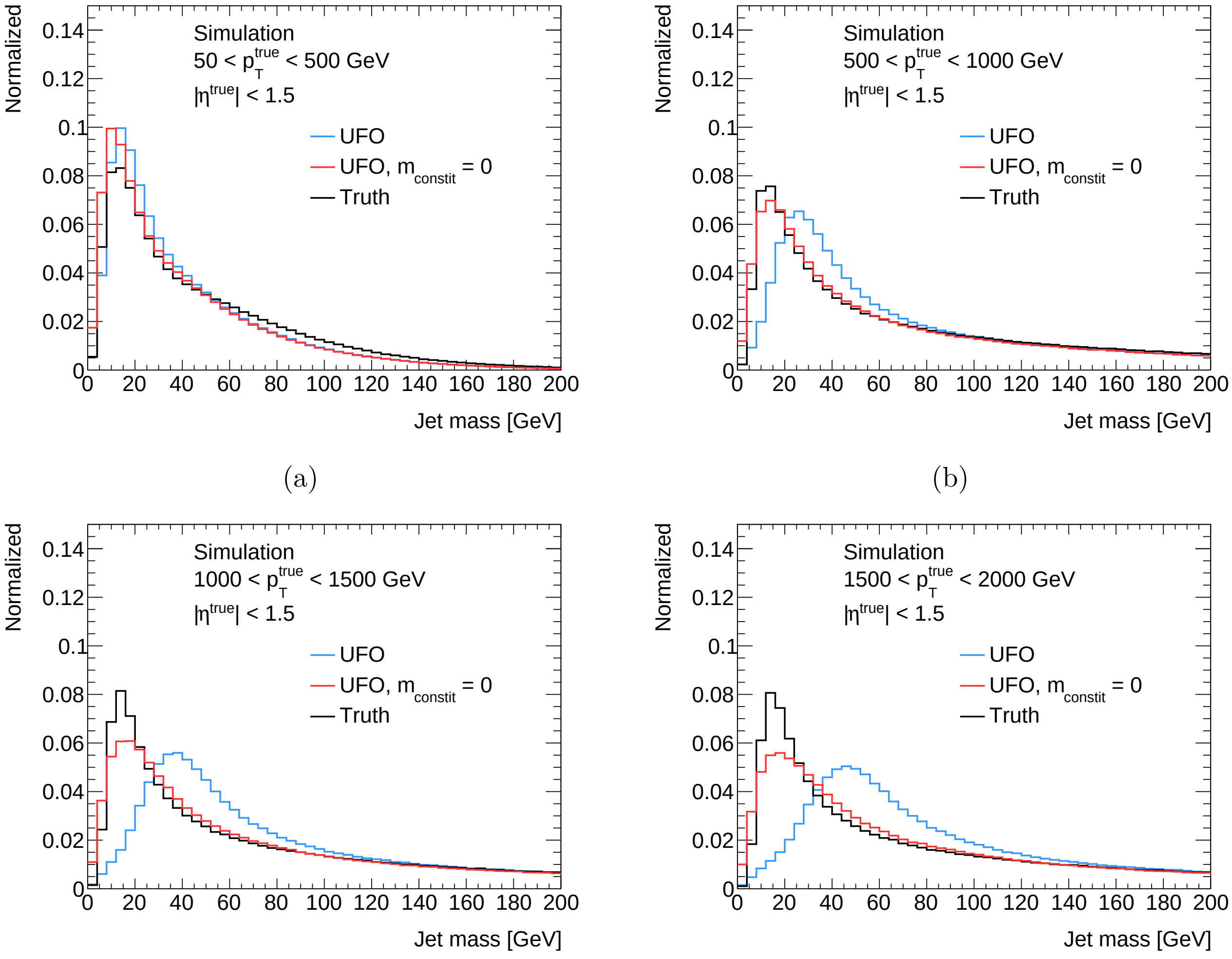
$$e_2 = \frac{\text{ECF2}}{(\text{ECF1})^2}, \tag{4.24}$$

$$e_3 = \frac{\text{ECF3}}{(\text{ECF1})^3}. \tag{4.25}$$

Finally, D_2 is defined as $e_3/(e_2)^3$ and similarly as τ_{21}, it is useful to identify two-body structures within jets.

The distributions of the jet mass, Q_w, τ_{21}, τ_{32} and D_2 for standard UFO jets, UFO jets built from massless constituents and truth jets in several bins of p_T^{true} are presented in Figures 4.26, 4.27, 4.28, 4.29 and 4.30 respectively.

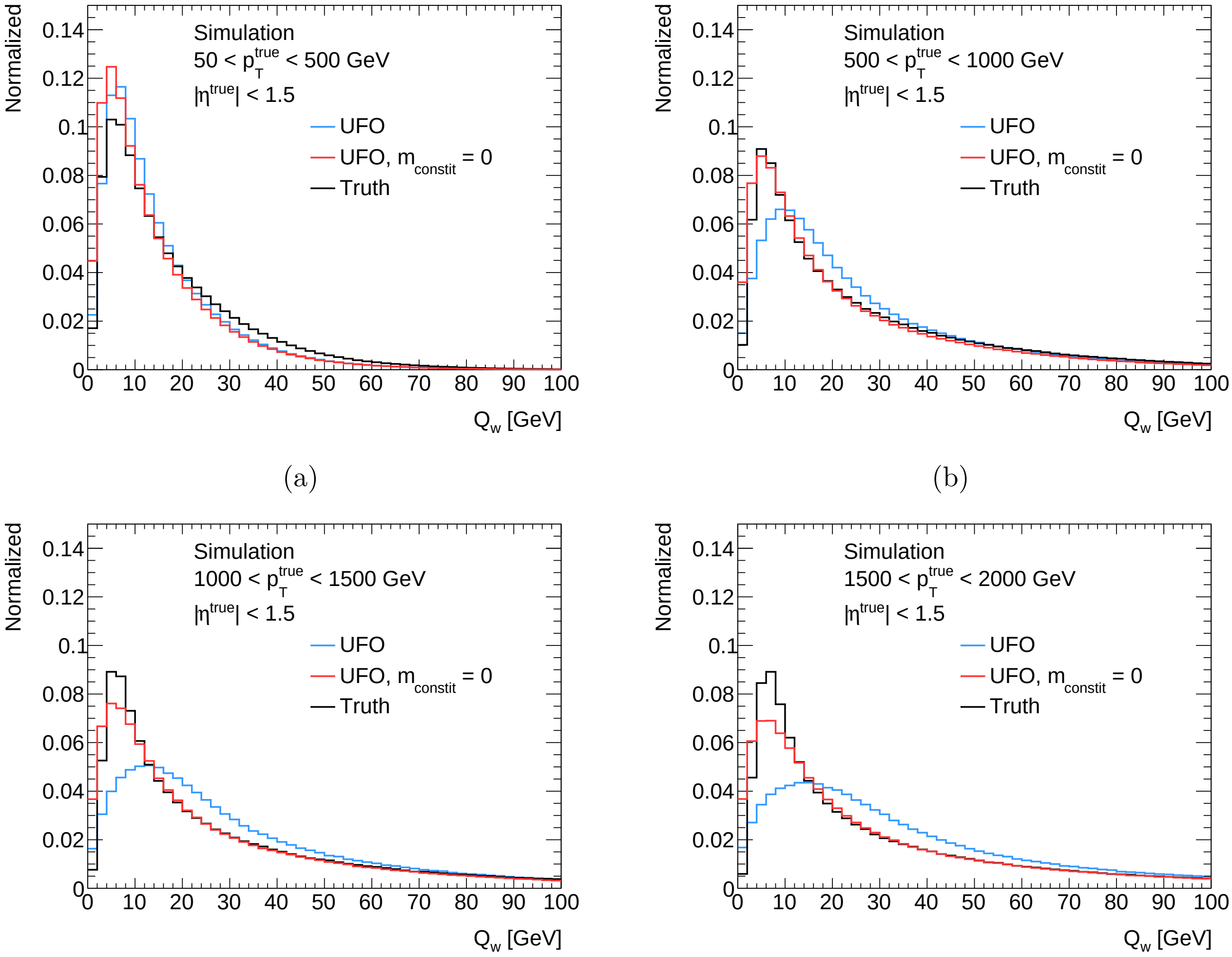
Notable differences can be observed for the jet mass and Q_w distributions, especially at high p_T. Indeed, the distributions of these variables obtained with UFO jets built from massless constituents are significantly much closer to the truth distributions than the standard UFO jet distributions. These variables are thus better reconstructed by configuring massless constituents instead of allowing important values for the UFO constituent mass. As expected, no significant differences can be observed between standard UFO jets and the ones composed of massless constituents considering the distributions of τ_{21}, τ_{32} and D_2, because these variables do not involve the constituent mass.



(a) (b)

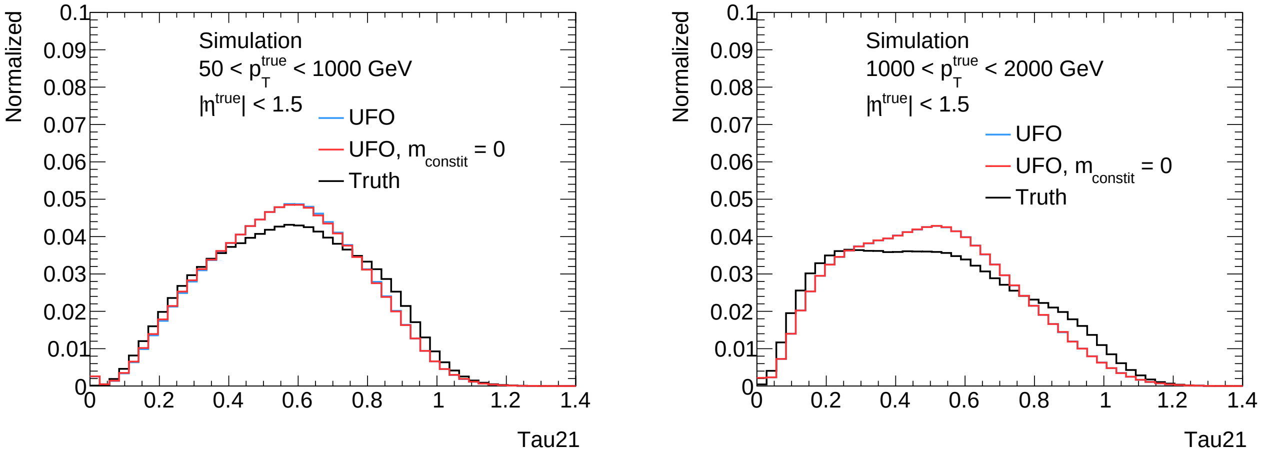
(c) (d)

Figure 4.26: Normalized distributions of the jet mass for UFO (both standard (blue) and with massless constituents (red)) and truth jets (black) located in the central part of the detector in different p_T^{true} bins.



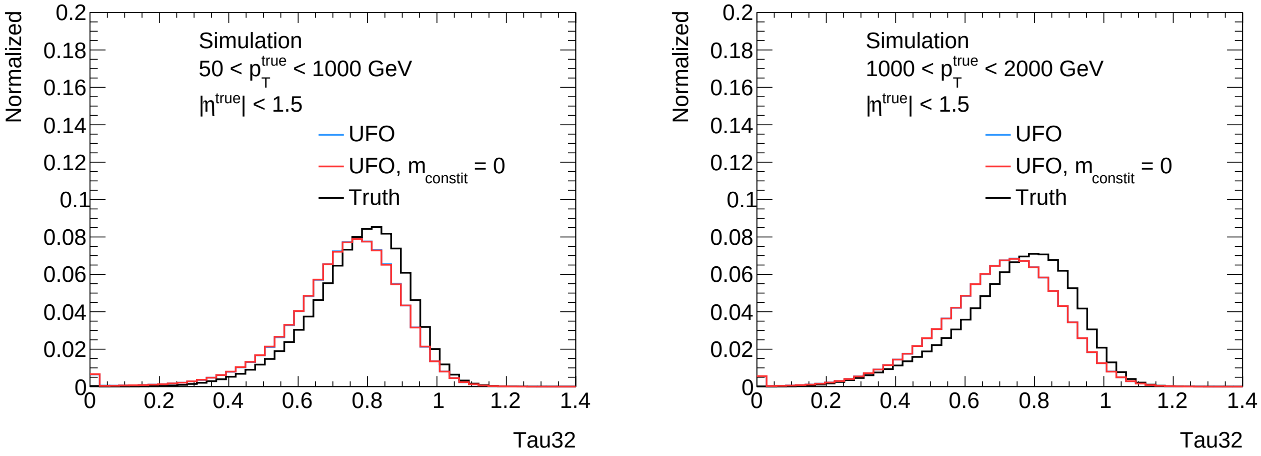
(a) (b) (c) (d)

Figure 4.27: Normalized distributions of Q_w for UFO (both standard and with massless constituents) and truth jets located in the central part of the detector in different p_T^{true} bins.



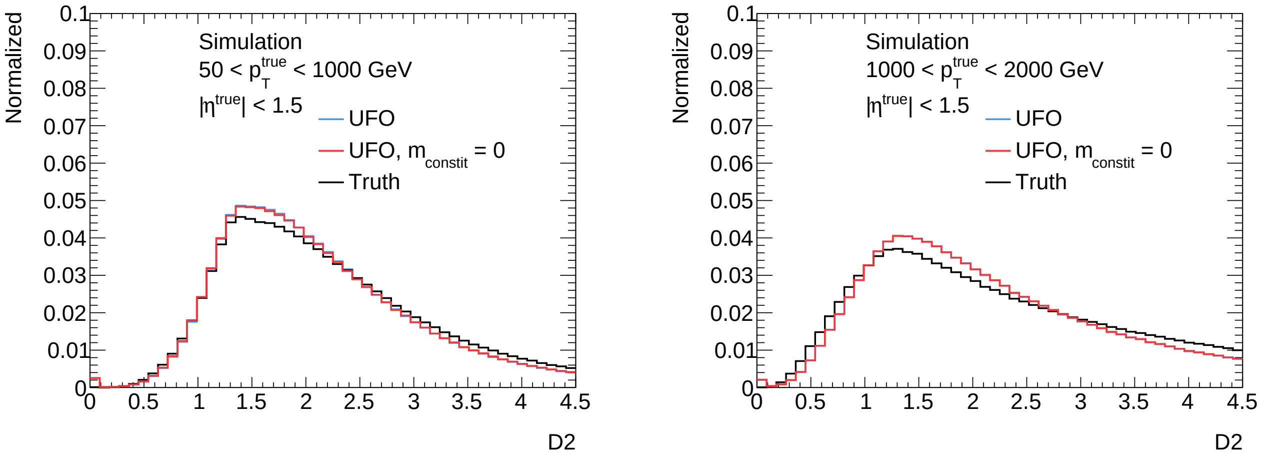
(a) (b)

Figure 4.28: Normalized distributions of τ_{21} for UFO (both standard and with massless constituents) and truth jets located in the central part of the detector in different p_T^{true} bins.



(a) (b)

Figure 4.29: Normalized distributions of τ_{32} for UFO (both standard and with massless constituents) and truth jets located in the central part of the detector in different p_T^{true} bins.



(a) (b)

Figure 4.30: Normalized distributions of D_2 for UFO (both standard and with massless constituents) and truth jets located in the central part of the detector in different p_T^{true} bins.

4.5.7 Conclusion

Two modifications in the UFO algorithm have been considered and studied.

The first one is a modification of the matching condition between tracks and clusters that consists in $\Delta R < \Delta R_{\text{cut}}$, with ΔR the angular distance between the track and the cluster. Instead of considering a value of ΔR_{cut} that depends on the cluster/track energy and η coordinate, a fixed value of ΔR_{cut} set to 0.05 allows to considerably improve the UFO small-R jet reconstruction in terms of angular, energy and mass resolutions, particularly in the forward region of the detector.

The second one consists in adjusting the momentum of the UFO constituents in order to obtain massless constituents having the same energy and angular coordinates. This modification allows notably to improve the reconstruction of the mass for UFO large-R jets by increasing the agreement with the truth jets, especially at high momentum.

These modifications in the UFO algorithm have thus been implemented in the ATLAS software framework for future jet reconstruction, and these studies allowed me to become an author of the ATLAS collaboration.

Chapter 5

Search for emerging jets with ATLAS using LHC Run-3 data

Contents

This section will present my contribution to the ATLAS Run 3 emerging jets analysis that led to an article [57] that has been submitted to the journal «Reports on Progress in Physics». I notably contributed to the determination of the event selection, the background estimation, the evaluation of the systematic uncertainties and the statistical interpretation for one of the four channels of the analysis. This section will present all of these elements and end by a presentation of the final analysis results that include the ones obtained from the other channels as well.

5.1 Emerging jets

As explained in section 2.4, emerging jets can arise in theories of a strongly interacting QCD-like dark sector that is connected with the visible one through a new interaction. At hadron colliders, dark quarks may be produced and initiate the production of dark jets containing dark hadrons. By considering the case where the dark hadrons are unstable and decay back to SM particles with macroscopic decay lengths (with respect to the detector size), the produced signature corresponds to visible signals emerging gradually in the detector.

To produce such a signature, it is necessary to require a large hierarchy between the mass of the mediator that connects the visible and dark sectors, and the mass scale of the dark sector [58]. Otherwise, the experimental signature consists in low-momentum and isotropically distributed particles that is generally named «soft unclustered energy patterns». For the model considered in this analysis, the mass of the mediator is considered to be at the TeV scale, while the one of the dark sector is at the GeV scale. In this way, the energy of the dark quarks coming from the decay of the mediator will be larger than the confinement scale of the dark gauge group, allowing them to shower and hadronize into dark hadrons. Additionally, by considering the dark hadrons decaying to SM quarks with a decay length in the order of millimeters, the dark jets will progressively appear in the detector as schematized in Figure 5.1 [58].

In this analysis, a new heavy Z' mediator is considered to act as a portal between the visible and the dark sectors. It couples to both SM quarks q and dark quarks q_d with coupling constants noted g_q and g_{q_d} respectively. The production process is illustrated in Figure 5.2 [57] : the Z' is produced in pp collisions via quark - anti-quark annihilation ($q\bar{q} \to Z'$), before decaying into a pair of dark quarks ($Z' \to q_d\bar{q}_d$). Each of these dark quarks leads to a dark jet that is considered to contain only dark mesons : π_d and ρ_d. The dark baryon production is not considered since it is suppressed by a factor of $1/N_{C_d}^2$, where N_{C_d} is the number of dark colors [59]. The dark rho mesons are set to promptly decay via $\rho_d \to \pi_d\pi_d$ while the dark pions are forced to decay to down quarks. Hence, only the dark pion life-time τ_{π_d} is considered to parametrize the dark particles decay length.

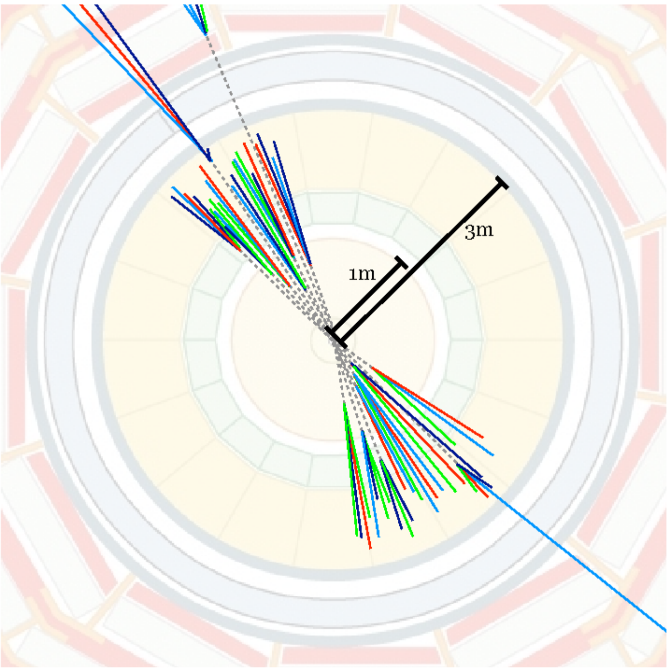


Figure 5.1: Schematic representation of a pair production of emerging jets in the transverse plane of a detector. The dashed lines correspond to long-lived dark mesons, while solid colored lines represent SM particles [58].

Following the model described in [58], the number of dark colors N_{C_d} is set to 3 and the number of dark quark flavors N_{f_d} to 7. It allows to obtain a QCD-like phenomenology, and to be able to use perturbative theory for computation. Moreover, the dark sector confinement scale Λ_d, the dark quark mass m_{q_d}, the dark pion mass m_{π_d} and the dark rho mass m_{ρ_d} satisfy the relation $m_{\rho_d} = 2\Lambda_d = 2m_{q_d} = 4m_{\pi_d}$. Three different sets of parameters for the dark particle masses are considered, corresponding to m_{π_d} values of 5, 10 and 20 GeV. Moreover, each signal is characterized by a Z' mass and a fixed proper decay length for the dark pions. The former ranges from 0.6 to 3.5 TeV while the latter ranges from 1 to 1000 mm in order to ensure that the majority of the dark pion decays occur in the inner detector, with a decay length in the detector frame given by $\gamma\beta c\tau_{\pi_d}$. The table 5.1 summarizes the parameter values of the model.

$(m_{\pi_d}, \Lambda_d, m_{\rho_d})$ [GeV]	(5, 10, 20), (10, 20, 40), (20, 40, 80)
$m_{Z'}$ [TeV]	**0.6**, 0.8, 1.0, 1.2, **1.5**, 1.8, 2.2, 2.6, **3.0**, 3.5
$c\tau_{\pi_d}$ [mm]	1, **5**, 10, **50**, 100, 500, 1000

Table 5.1: Parameters of the benchmark models used in the analysis. The mediator masses and dark pion proper decay lengths that have been exploited during the discriminating studies are indicated in bold, while the other samples have been generated subsequently.

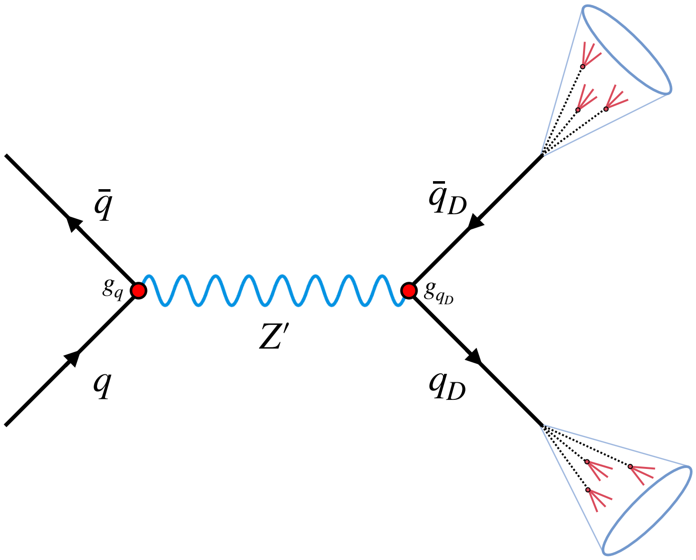


Figure 5.2: Diagram illustrating the production of a pair of emerging jets via an s-channel mediator Z'. The emerging jets are represented as blue cones, the dashed lines correspond to the dark pions and the red lines represent their SM decay products [57].

5.2 Data and MC samples

5.2.1 Data

This analysis uses 51.8 fb^{-1} of data from pp collisions at $\sqrt{s} = 13.6$ TeV that have been collected by ATLAS in 2022 and 2023. To be considered, data events must have been collected under stable beam conditions with all the detector sub-components being functional during the data acquisition. Moreover, additional data quality requirements are applied to exploit only data that allows reliable reconstruction [60]. For instance, events that are compatible with noise burst in the LAr calorimeters are rejected.

5.2.2 MC simulations

MC simulated background and signal events are used to study the specific properties of the signal in order to determine the discriminating variables that will be used in the analysis event selection. MC signal events are also used to compute the signal acceptance and the associated systematic uncertainties that are both used during the statistical interpretation.

Background

The final state corresponding to the emerging jet signal is a pair of two energetic jets. As a consequence, the main background that is considered corresponds to di-jet events containing two QCD jets that are initiated by gluons or quarks. Multi-jet production has been generated using Pythia 8.230 [30] with Matrix Elements (ME) at LO for di-jet production which has been matched to the parton shower. The NNPDF2.3LO parton distribution functions collection [54] was used in the ME generation, the parton shower, and the simulation of the multi-parton interactions and the A14 set of tuned parameters has been used [55].

The process corresponding to the production of a top quark associated with an anti-top quark (referred to as $t\bar{t}$) also constitutes a source of background that has been considered, even if it is less important than the di-jet events. Each top quark decays via $t \rightarrow Wb$, while the W boson can decay via $W \rightarrow q\bar{q}'$ (hadronic decay) or $W \rightarrow l\bar{\nu}$ (leptonic decay). The $t\bar{t}$ process can thus produce either two hadronic decays, one hadronic and one leptonic decay or two leptonic decays. The $t\bar{t}$ production has been modelled using the Powheg Box v2 [61, 62, 63, 64] generator at NLO with the NNPDF3.0NLO parton distribution functions collection [65]. The events have been interfaced to Pythia 8.230 to model the parton shower, hadronization and underlying event with the NNPDF2.3LO set of parton distribution functions and using the A14 set of parameters. The decays of bottom and charm hadrons have been executed by EvtGen 1.6.0 [66].

Moreover, in order to study the efficiencies of the triggers that are used in the analysis (cf section 5.4) and to compare the distributions of jet kinematic variables between data and MC events, a control region corresponding to single-muon events is considered. Additional processes that produce both muon and jets are thus simulated :

- $W \rightarrow \mu\bar{\nu}_{\mu}+$ jets,
- $W \rightarrow \tau\bar{\nu}_{\tau}+$ jets, with $\tau \rightarrow \mu\bar{\nu}_{\mu}\nu_{\tau}$.

These W + jets samples have been generated with Sherpa 2.2.14 [31] using Matrix Elements at NLO for up to two partons and LO for up to five partons calculated with the Comix [67] and OpenLoops [68, 69, 70] libraries. They were matched with the Sherpa parton shower [71] using the MEPS@NLO prescription [72, 73, 74, 75] and the set of tuned parameters developed by the Sherpa authors. The NNPDF3.0NNLO set of PDFs has been used.

Signal

The signal samples corresponding to the process $pp \rightarrow Z^{'} \rightarrow q_d\bar{q}_d$ have been simulated using the Hidden Valley module [76, 77] of Pythia 8.309 [78], and by configuring the dark QCD parameter values and the dark mesons decays such as described in section 5.1. The different dark pions masses allow to cover variations within the dark hadronization process. The $Z^{'}$ branching ratios are taken as BR($Z^{'} \rightarrow q\bar{q}$) = 0.001 and BR($Z^{'} \rightarrow q_d\bar{q}_d$) = 0.999 in order to maximize the decay to the desired final state. These branching ratios set the values of the coupling constants of the Z' to quarks and dark quarks g_q and g_{q_d}, but the samples can be re-scaled.

All simulated events are processed through a simulation of the ATLAS detector using Geant4 [34], and the effect of pile-up is modelled by overlaying the simulated hard-scattering process with inelastic pp collisions generated from EPOS 2.0.1.4 [79] and Pythia 8.308.

In order to be compared to data events, each MC event is normalized according to a total weight given by :

$$w = \sigma \times \mathcal{L}_{\text{int}} \times f_{\text{eff}} \times k \times (w_{\text{MC}} / \sum w_{\text{MC}}) \times w_{\text{pile-up}} \tag{5.1}$$

where σ is the cross-section of the process as computed by the generator, $\mathcal{L}_{\text{int}}$ is the integrated luminosity of the data to which the MC events are compared, f_{eff} is the efficiency of the event filter of the generator, k is a factor that takes into account higher order terms in the cross-section computation, w_{MC} is the MC event weight provided by the particle generator, $\sum w_{\text{MC}}$ is the sum of all the MC event weights in a given sample and $w_{\text{pile-up}}$ is a weight that is applied such as the distribution of μ (as described in 3.3.2) is the same between MC and data events.

5.3 Event reconstruction

5.3.1 Jets

As described in sections 4.3 and 4.4, there are different ways to reconstruct jets in ATLAS, depending on the input constituents or on the jet algorithm parameters that are used. The reconstruction method in this analysis has been determined by considering the particular topology of emerging jets. Indeed, each emerging jet undergoes a process of double hadronization : a first one in the dark sector, and a second one in the SM visible sector, resulting in wide opening angles between the final hadrons. It is thus necessary to use large-R jets with $R = 1.0$ instead of small-R jets to capture most of the produced hadronic shower.

Moreover, emerging jets are likely to be associated to a large amount of displaced tracks that do not originate from the primary hard-scattering vertex. As a consequence, constituents obtained from the combination of tracks and topo-clusters such as PFlow or UFO cannot be used in this analysis, because their reconstruction considers only tracks that are matched to the primary hard-scattering vertex. Consequently, topological clusters at the EM scale (EMTopo) are chosen as jet constituents.

The procedure of jet reconstruction is the following : the input constituents are first clustered into jet with $R = 0.4$, and these small-R jets are then reclustered into $R = 1.0$ jets. Both clustering are done with the anti-k_t algorithm. The small-R jet four-momenta are calibrated with the techniques described in 4.4.3. For each large-R jet, the calibrated four-momentum is obtained by summing the four-momentum of its calibrated small-R jet constituents, while the uncertainties from the small-R jet calibration are propagated to the large-R jet.

This way to proceed has physical and technical benefits. First, the minimum p_T threshold that is applied to small-R jets functions as a grooming procedure similar to the trimming or soft-drop one, by removing contaminations from soft particles. Then, small-R EMTopo jets are commonly used in other ATLAS analyses that are searching for LLPs, and thus the same calibration can be used. For example, an alternative could have been to consider large-R jets built directly from LCTopo constituents, but this jet collection is no longer supported in Run-3 by the ATLAS jet performance group, and no calibration nor uncertainties are available.

The selections applied to the uncalibrated small-R and large-R jets are presented in Tables 5.2.

	Small-R	**Large-R**
p_T	> 15 GeV	> 200 GeV
$\lvert\eta\rvert$	< 2.5	< 1.8

Table 5.2: Selection criteria applied to the jets used in the analysis.

5.3.2 Tracks and displaced vertices

The tracks are reconstructed according to the procedure described in section 4.1 with the parameters shown in Figure 4.3. Both Primary and Large Radius Tracking algorithms are used, and standard selections are applied to the reconstructed tracks as summarized in Table 5.3.

	Primary	**Large Radius Tracking**
p_T	> 0.5 GeV	> 1 GeV
$\lvert\eta\rvert$	< 2.7	< 3

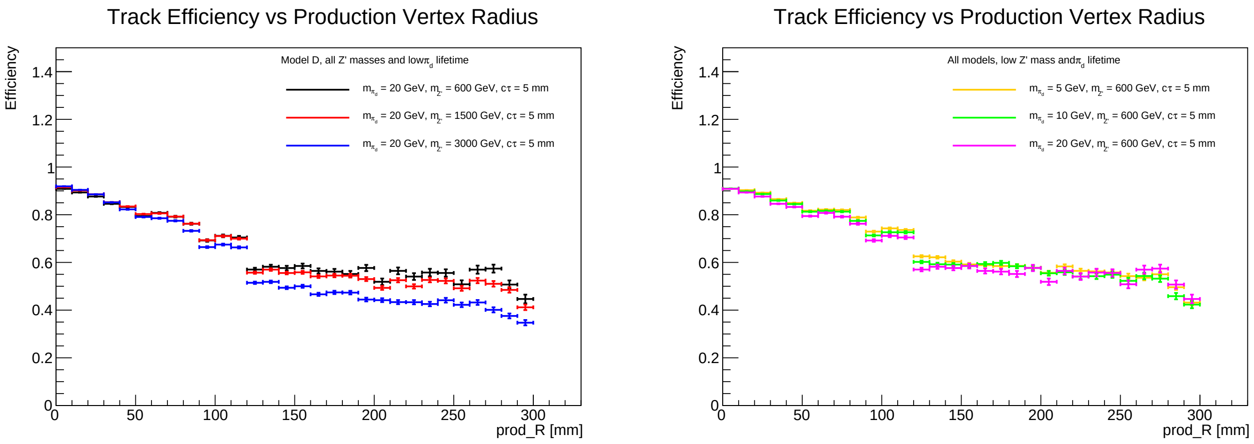
Table 5.3: Selection criteria applied to the tracks used in the analysis.

The track reconstruction efficiency, defined as the fraction of truth charged particles that are matched to a reconstructed track, evaluated in the signal samples as a function of the particle production radius is shown in Figure 5.3. It is possible to see that at high production radius (> 120 mm), the efficiency for the signal corresponding to $m_{Z'} = 3$ TeV is degraded with respect to lower $m_{Z'}$ values. Indeed, at higher $m_{Z'}$, the produced charged particles will be more collimated, resulting in tracks that may more probably share hits with the others, and thus be rejected by the tracking algorithm. On the other hand, no differences in terms of efficiency is observed between signals corresponding to different m_{π_d} values.

In order to compute track-based jet observables, the ghost association technique has been used to associate the tracks to the jets, by considering each track as a ghost and proceeding as described in section 4.4.1.

The displaced vertices are reconstructed according to the procedure described in section 4.2 from the combined collection of standard and large-radius tracks. The secondary vertex reconstruction efficiency evaluated in the signal samples as a function of the decay radius of the truth dark pions is shown in Figure 5.4. This efficiency is evaluated as the fraction of individual dark pion decays that are matched to a reconstructed vertex with the matching procedure described in [40].

While the reconstruction efficiency is similar between the samples that differ only in the Z' mass value, disparities can be observed between samples having different m_{π_d} values. Indeed, at a fixed dark pion decay radius, the vertex reconstruction efficiency increases with the mass of the dark



(a) (b)

Figure 5.3: Track reconstruction efficiency as a function of the production radius for signal samples corresponding to (a) different $m_{Z'}$ and fixed m_{π_d} and $c\tau_{\pi_d}$ and (b) different m_{π_d} and fixed $m_{Z'}$ and $c\tau_{\pi_d}$.

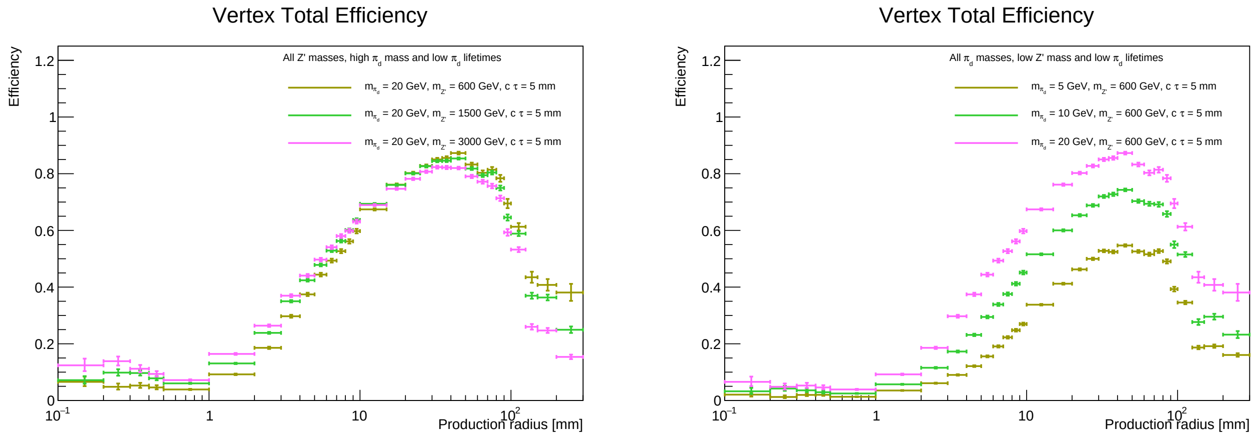
pion for the same reasons as described in section 4.2.2 : the efficiency increases with the vertex track multiplicity, while the latter increases with the dark pion mass.

Additional requirements are applied to the reconstructed secondary vertices in order to remove vertices coming from SM particle decays.

The mass of the vertex is required to be higher than 0.6 GeV : indeed, it allows to reject vertices coming from the decay of kaons notably, which have a mass about 0.5 GeV (cf Figure 5.5.a). The number of tracks associated to a vertex is required to be strictly greater than 2, in order to remove vertices that can come from photons conversions producing electron - positron pairs (cf Figure 5.5.b).

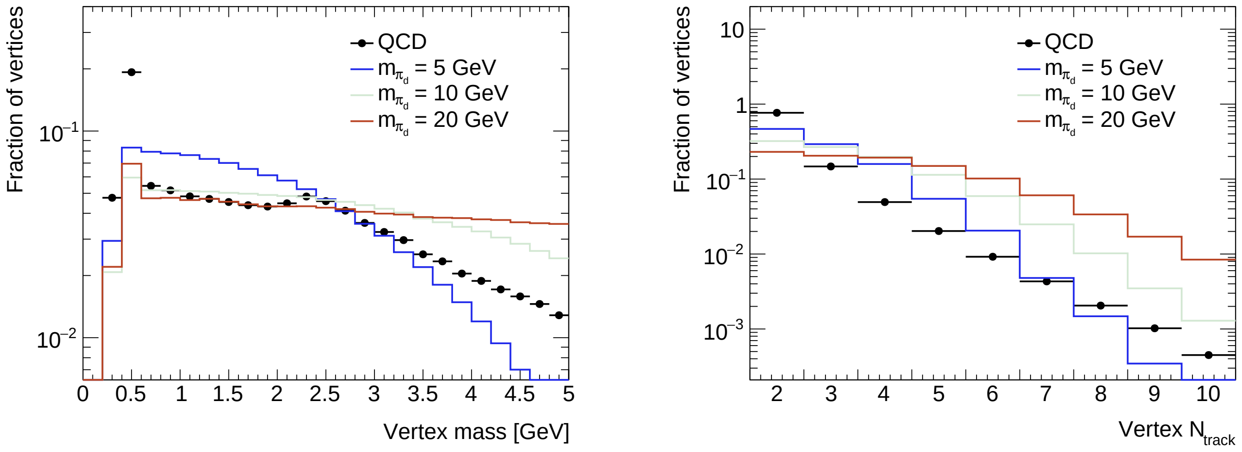
Moreover, displaced vertices can be produced via the interactions of SM particles with the detector material. The vertices are thus required to pass a material map veto, that reject secondary vertices having a position consistent with the position of the detector material. This veto may potentially reject signal vertices but its application is crucial in this analysis. Indeed, Figure 5.6.a shows the distribution of the reconstructed secondary vertices from QCD background events in the transverse plane of the detector (x, y), which is highly correlated to the position of the material in the pixel detector. Figure 5.6.b shows the distribution of the remaining vertices after the veto being applied.

Finally, each vertex satisfying these requirements is geometrically matched to a reconstructed jet if passing the criteria $\Delta R < 1.0$, with ΔR the angular distance between the vertex and the jet.



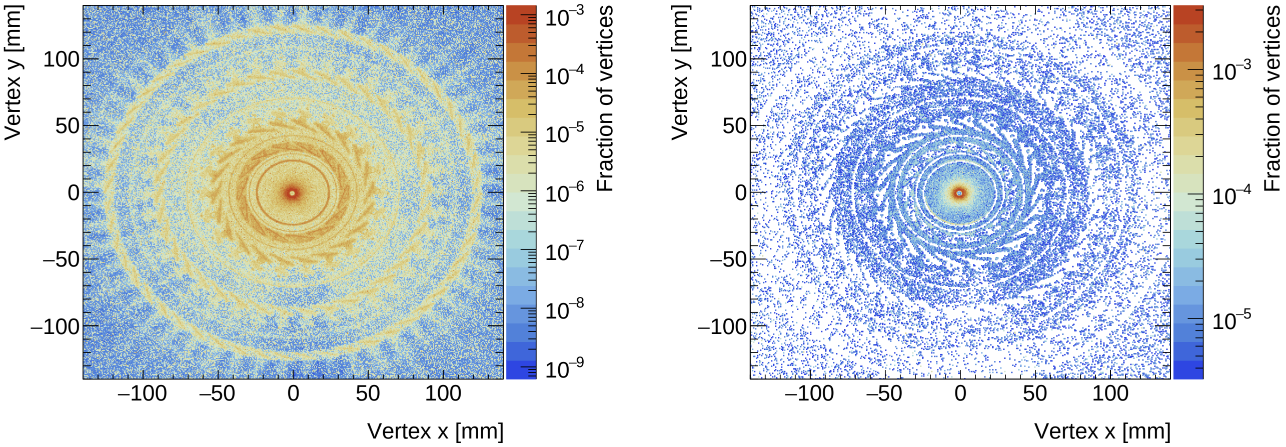
(a) (b)

Figure 5.4: Secondary vertex reconstruction efficiency as a function of the production radius for signal samples corresponding to (a) different $m_{Z'}$ and fixed m_{π_d} and $c\tau_{\pi_d}$ and (b) different m_{π_d} and fixed $m_{Z'}$ and $c\tau_{\pi_d}$.



(a) (b)

Figure 5.5: Normalized distributions of (a) the vertex mass and (b) the number of associated tracks to a secondary vertex for QCD and signal events corresponding to $m_{Z'} = 1.5$ TeV and $c\tau_{\pi_d} = 50$ mm after the high-p_T jet trigger being applied (defined in section 5.4.1).



(a) (b)

Figure 5.6: Normalized distributions of the reconstructed secondary vertices from QCD background events in the transverse plane (x, y) (a) without any selection and (b) after the material map veto being applied.

5.4 Pre-selections

First, the events are required to have one reconstructed primary vertex associated to at least two tracks. As mentioned in section 4.2.1, the hard-scattering primary vertex is selected as the one having the greatest $\sum p_T^2$, where the sum occurs on the primary tracks associated to the vertex.

Triggers The events are selected using one of the two jet-based triggers that are considered in this analysis.

The first one selects events having at least one large-R jet reconstructed in the same way as described in section 5.3.1 and that satisfy $p_T > 460$ GeV : it is referred as the «high-p_T jet trigger».

The second one referred as the «emerging jet trigger» selects events with at least one large-R PFlow jet reconstructed with the anti-k_t algorithm and with the CS+SK and Soft-Drop procedures applied. This jet must satisfy $p_T > 200$ GeV, $|\eta| < 1.8$ and PTF < 0.08, where PTF stands for «Prompt Track Fraction» and is given by :

$$\mathrm{PTF} = \frac{\sum p_T^{\mathrm{trk}}}{p_T^{\mathrm{jet}}} \tag{5.2}$$

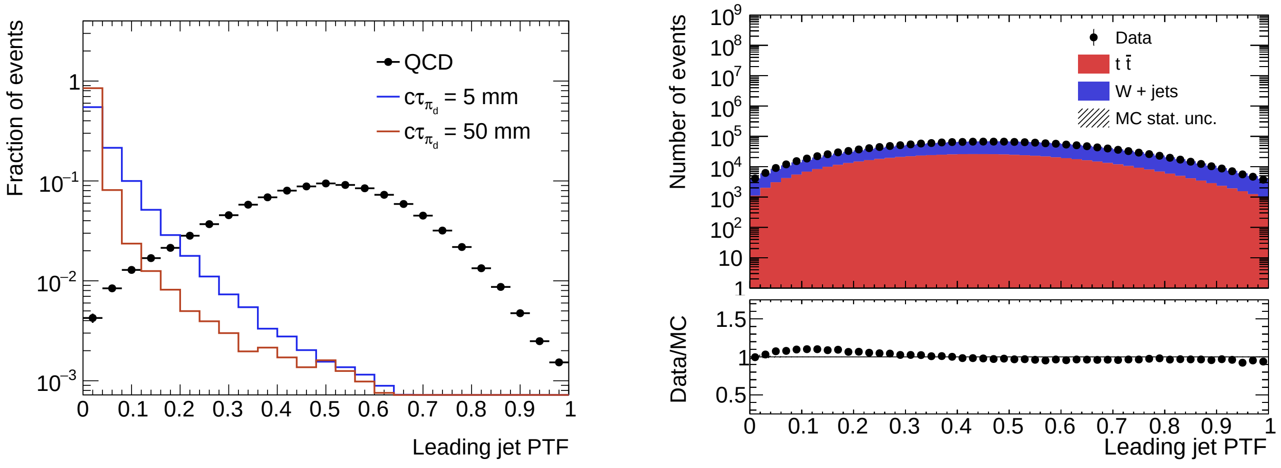
with the sum occurring on all the tracks within $\Delta R < 1.2$ of the jet that satisfy : $p_T > 1$ GeV, $|d_0/\sigma_{d_0}| < 2.5$ with σ_{d_0} the uncertainty on the measurement of the track d_0, and $|z_{\mathrm{PV}} - z_0| < 10$ mm with z_{PV} the position of the hard-scattering primary vertex along the beam axis. The selection on d_0 allows to select prompt tracks originating from the hard-scattering primary vertex while the one on z_0 aims to reduce contributions from pile-up. Emerging jets are expected to be associated mainly to displaced tracks and thus to have low PTF values as shown in Figure 5.7.a. The PTF distributions in data and MC events in the muon control region (defined below) is shown in Figure 5.7.b and present a good agreement within around 10%.

Trigger efficiency The efficiencies of these two triggers have been studied with data and MC events in a single-muon control region where signal events are not selected. This region is defined using a single-muon trigger that select events by requiring at least one muon with $p_T > 24$ GeV at the HLT level. In addition, it is required that the reconstructed muon satisfy $p_T > 27$ GeV and pass an isolation criteria, ensuring that the muon used in the trigger decision was not produced within a jet. Moreover, the events are required to have at least one large-R jet reconstructed with the procedure described in 5.3.1.

The efficiency of the high-p_T (emerging) jet trigger is evaluated in this muon control region as :

$$\epsilon = \frac{\text{Number of event selected by the high-}p_T\text{ (emerging) jet and muon triggers}}{\text{Number of event selected by the muon trigger}}. \tag{5.3}$$

It has been evaluated both in data and MC samples (W + jets and $t\bar{t}$) as a function of the leading jet p_T, and also as a function of the leading jet PTF for the emerging jet trigger as shown in Figures 5.8 and 5.9.



(a) (b)

Figure 5.7: PTF distributions (a) for QCD and signal events corresponding to $m_{Z'} = 1.5$ TeV and $m_{\pi_d} = 10$ GeV after the application of the high-p_T jet trigger and (b) for data and MC events in the muon control region.

A fit to the efficiency in data and MC has been done to identify the threshold in p_T for the high-p_T and emerging jet triggers, and in PTF for the emerging jet trigger, at which the efficiency reaches a constant value. For the high-p_T jet trigger, 98 % of the maximum efficiency is obtained for jets that satisfy $p_T > 520$ GeV. Moreover, this trigger is very well modelled in MC simulations as shown in Figure 5.8 where the data and MC efficiency curves are in excellent agreement. The fact that the high-p_T jet trigger is not fully efficient until 520 GeV while the p_T threshold that defines the trigger is 460 GeV can be explained by considering that the jets that are used for trigger decisions are different from the jets reconstructed with the offline software.

For the emerging jet trigger, the maximum efficiency is obtained for jets passing $p_T > 300$ GeV and PTF < 0.04. In order to increase the sensitivity to lower $m_{Z'}$ signal samples, the turn-on region has been exploited by reducing the offline threshold on p_T to 250 GeV. Moreover, as shown in Figures 5.9.a and 5.9.b, the trigger efficiency curves are different between data and MC events unlike the high p_T jet trigger ones. Therefore, a scale factor is applied to the MC event weights in order to correct it. This scale factor depends on the p_T of the leading jet and corresponds to $\epsilon_{\text{data}}(p_T)/\epsilon_{\text{MC}}(p_T)$, with ϵ_{data} and ϵ_{MC} the efficiency of the emerging jet trigger at a given p_T obtained from the fit function for data and MC respectively. The value of this scale factor ranges from 0.72 at $p_T = 250$ GeV to 0.96 at $p_T > 300$ GeV. Even if the emerging jet trigger is less efficient to select events with respect to the high-p_T jet trigger, it allows to select signal events down to $m_{Z'} = 0.6$ GeV, which is the lowest $Z^{'}$ mass value that is investigated in this analysis, and thus presents a significant interest.

The trigger efficiencies evaluated for the signal events as a function of the leading jet p_T are shown in Figure 5.10. No particular selections are applied in this case.

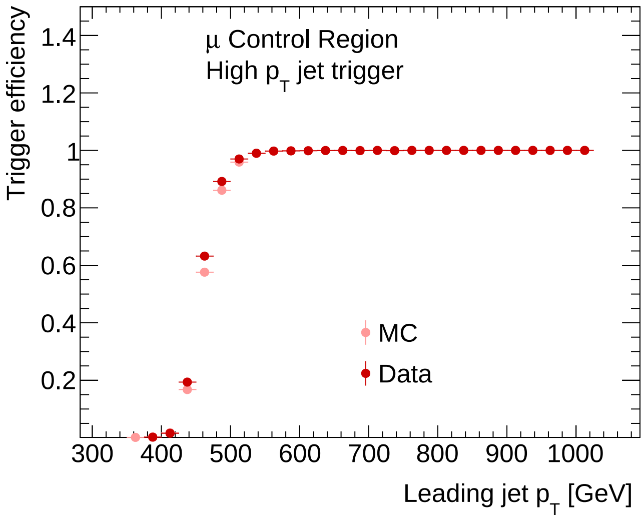
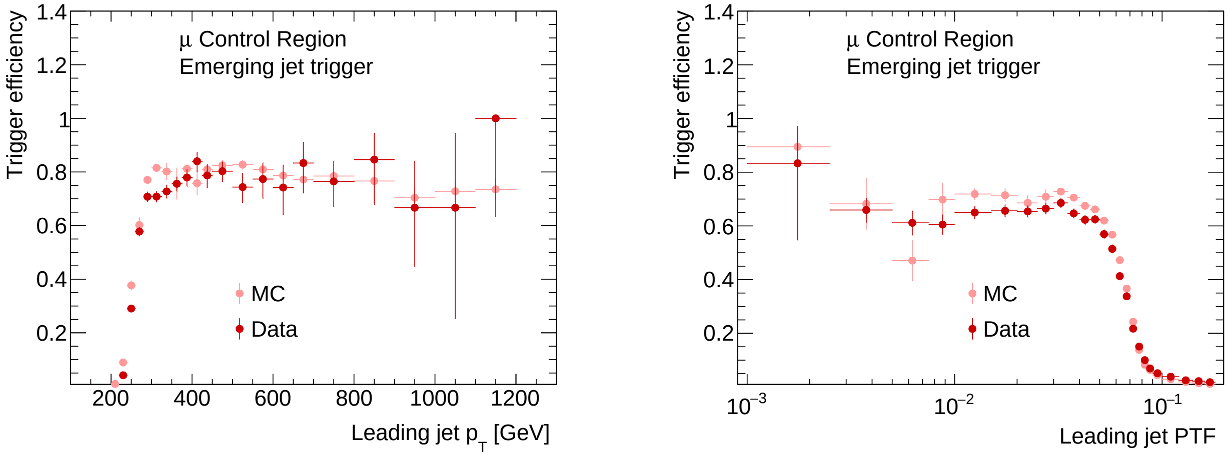


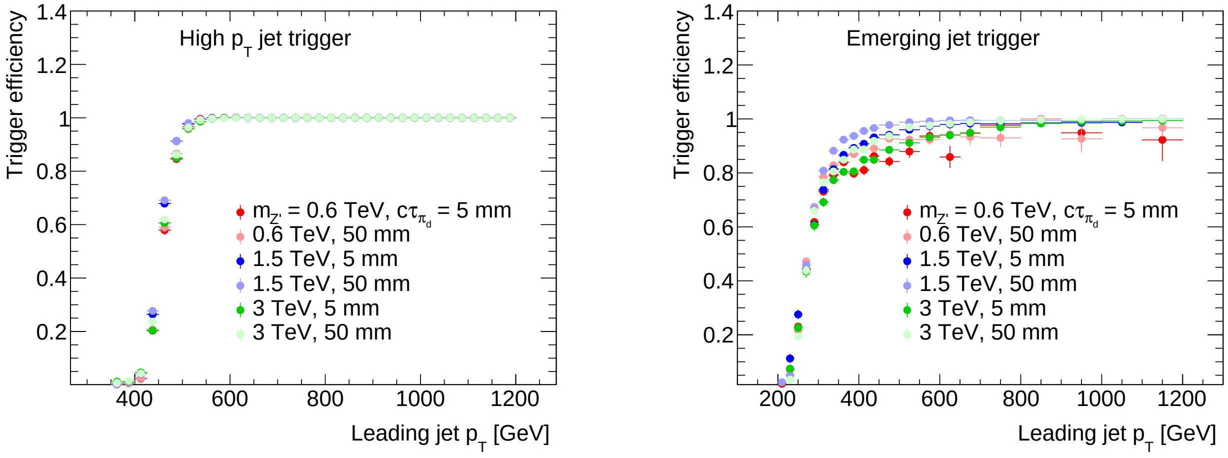
Figure 5.8: Efficiency of the high-p_T trigger as a function of the leading jet p_T evaluated for both data and MC events in the muon control region.



(a)

(b)

Figure 5.9: Efficiency of the emerging jet trigger as a function of (a) the leading jet p_T with $\mathrm{PTF}^{\mathrm{lead.}} < 0.04$ applied, and as a function of (b) the leading jet PTF with $p_T^{\mathrm{lead.}} > 250$ GeV applied, evaluated for both data and MC events in the muon control region.



(a)

(b)

Figure 5.10: Efficiency of the (a) high-p_T and (b) emerging jet triggers as a function of the leading jet p_T evaluated for signal events corresponding to $m_{\pi_d} = 10$ GeV.

5.5 Analysis strategy

The analysis is divided into two different approaches.

The first one, referred as «cut-based», consists in applying selections to jet observables that discriminate signal and background events, in order to select the former and maximally reject the latter. The determination of the potential discriminating variables has been done using MC simulated events. The whole selections allow to define a signal region (SR) expected to be highly sensitive to the emerging jet signature.

The second approach, referred as «Machine Learning (ML) based», uses a machine learning algorithm that has been trained to differentiate emerging jets from SM jets. In this approach, the SR is defined as containing events that have at least two jets identified as emerging jets.

While the ML-based approach is expected to maximize the sensitivity for the specific models that are investigated, the cut-based approach is more easily reinterpretable with alternative theoretical models. Indeed, the tagging algorithm used in the ML-based approach exploits low-level track information that come from the tracking algorithm and thus rely on a full simulation of the detector, while reinterpretation frameworks generally do not (cf section 6).

In both approaches, the analysis principle is similar : the expected background contribution from SM processes in the SR in terms of number of events is evaluated, before being compared to the observed number of events in data. If no excess is observed in data regarding the expected background, a statistical interpretation is performed to determine exclusion domains in the parameter space of the emerging jet theoretical model.

Finally, each approach is divided into a low and a high-m_{jj} region, each using one of the two triggers defined previously in association with a selection on the invariant mass of the two leading jets of the events, m_{jj}. The low-m_{jj} region is defined by selecting events with the emerging jet trigger and $m_{jj} \leq 1$ TeV, while the high-m_{jj} region selects events with the high-p_T jet trigger and $m_{jj} > 1$ TeV. The choice of 1 TeV is motivated by the fact that the high-p_T jet trigger mainly selects signal events roughly corresponding to $m_{Z'} > 1$ TeV because of the p_T requirement. In this way, the sensitivity to signal corresponding to $m_{Z'} \leq 1$ TeV (> 1 TeV) is maximized in the low(high)-m_{jj} region.

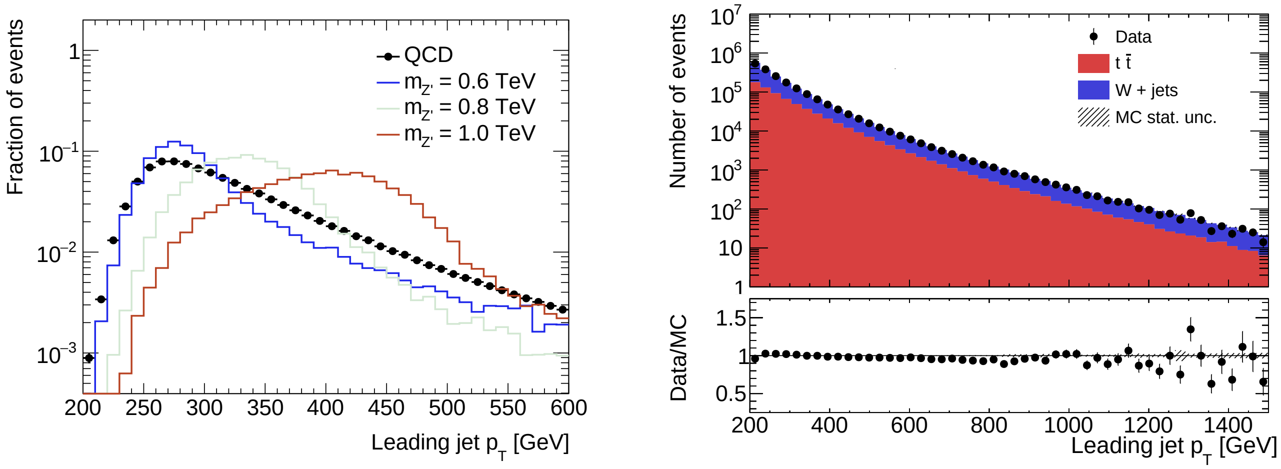
My contribution to this analysis is related to the low-m_{jj} region in the cut-based approach. I contributed to the determination of the event selection, the estimation of the background, the evaluation of the systematic uncertainties and the statistical interpretation which will be presented in the following sections.

5.6 Low-m_{jj} cut-based analysis

5.6.1 Pre-selections

Events are pre-selected with the emerging jet trigger. In addition, events must contain at least two reconstructed large-R jets satisfying $p_T > 200$ GeV, $|\eta| < 1.5$ and passing an overlap removal (OR) procedure with photon (i.e no photon can be found within $\Delta R = 1.0$ of the jets), in addition to the selection on the invariant mass of the two leading jets of the events $m_{jj} \leq 1$ TeV. The selection on η insures that the jet areas are within the inner detector acceptance, while the overlap removal procedure allows to reject photons that can produce low PTF jets and thus mimic the emerging jet signature.

The distributions of the leading jet p_T, η and m_{jj} for background and signal events are shown in Figures 5.11, 5.12 and 5.13, in addition to the corresponding data and MC distributions in the muon control region. Given the selection on m_{jj} that is applied, only the signal corresponding to $m_{Z'} = 0.6$ TeV is considered for the following studies.



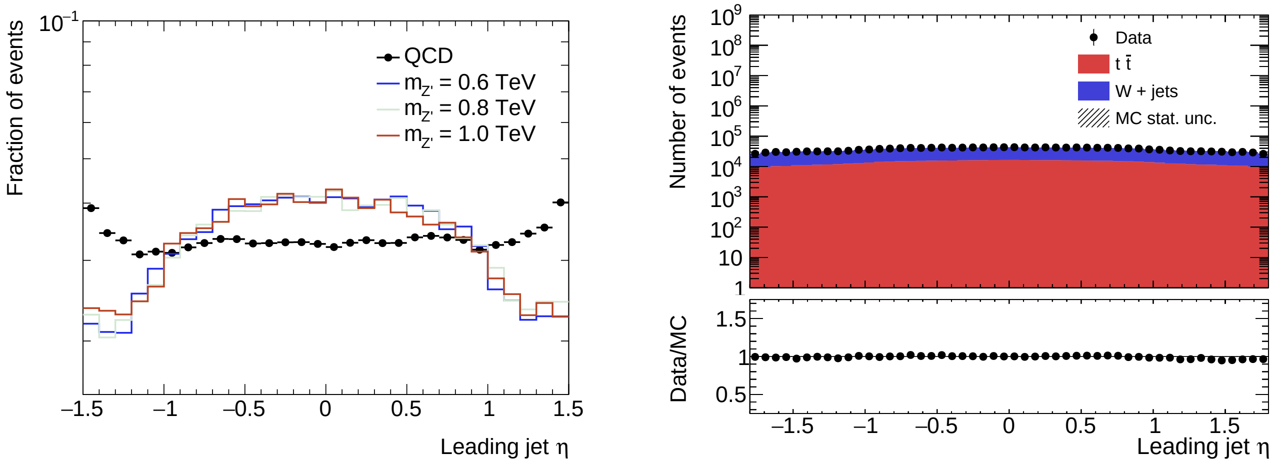
(a)

(b)

Figure 5.11: p_T distributions (a) for QCD and signal events corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm after the application of the emerging jet trigger and (b) for data and MC events in the muon control region.

As mentioned in section 5.4, selections are applied to the p_T and PTF of the reconstructed jet that is geometrically matched (i.e based on a ΔR requirement) to the jet that activate the emerging jet trigger at the HLT level. These selections correspond to $p_T > 250$ GeV and PTF < 0.04.

Moreover, for both leading and sub-leading jets, a selection is applied on a jet variable named ΔPTF, which is defined as $\Delta\text{PTF} = \text{PTF}(\text{no } \Delta z \text{ selection}) - \text{PTF}(\Delta z < 10 \text{ mm})$, with the PTF computed as given by (5.2). It corresponds to the difference between the PTF computed without any requirement on the z_0 coordinate of the tracks, and the PTF with the same selection as applied



(a)

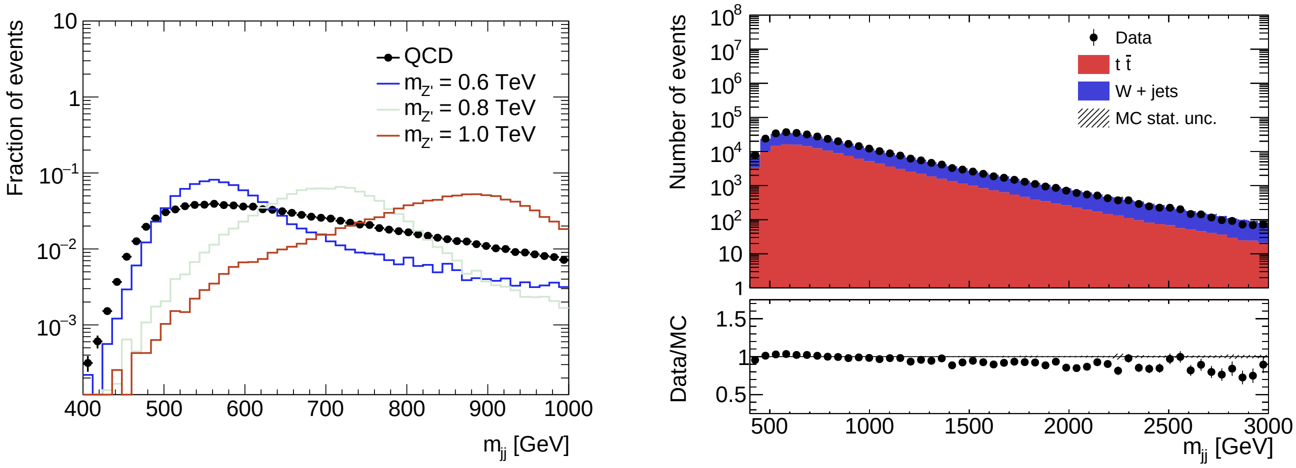
(b)

Figure 5.12: η distributions (a) for QCD and signal events corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm after the application of the emerging jet trigger and (b) for data and MC events in the muon control region.

in the trigger, i.e $\Delta z = |z_{\text{PV}} - z_0| < 10$ mm where z_{PV} is the coordinate of the hard-scattering primary vertex along the beam axis.

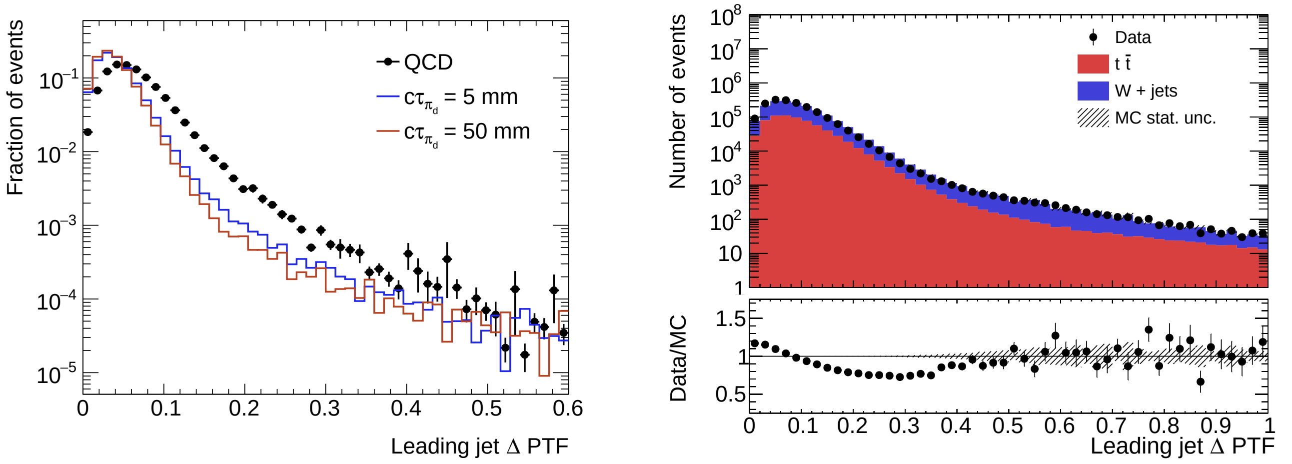
Indeed, the PTF of a jet as computed by the trigger relies on a correct determination of the hard-scattering primary vertex. Considering the possible case where a pile-up vertex is wrongly determined as the hard-scattering primary vertex, the PTF computation with the equation (5.2) may reject tracks that are however contained in the leading jets. It can thus possibly lead to an artificially low PTF value for jets in background events. Even if such a wrong designation of the hard-scattering primary vertex is expected to be rare, requiring a small ΔPTF for both leading and sub-leading jets, i.e requiring jets for which the PTF is not highly sensitive to the primary vertex position determination, can help to mitigate such effect. The selection that is applied corresponds to ΔPTF < 0.4. The distribution of the leading jet ΔPTF for background and signal events is shown in Figure 5.14 in addition to the data and MC comparison in the muon control region. The agreement between data and MC events is not as great as the one for the PTF, but the acceptance of a selection at 0.4 is similar between data and MC.

The pre-selections applied are summarized in Table 5.4.



(a) (b)

Figure 5.13: m_{jj} distributions (a) for QCD and signal events corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm after the application of the emerging jet trigger and (b) for data and MC events in the muon control region.



(a) (b)

Figure 5.14: ΔPTF distributions (a) for QCD and signal events corresponding to $m_{Z'} = 1.5$ TeV and $m_{\pi_d} = 10$ GeV after the application of the high-p_T jet trigger and (b) for data and MC events in the muon control region.

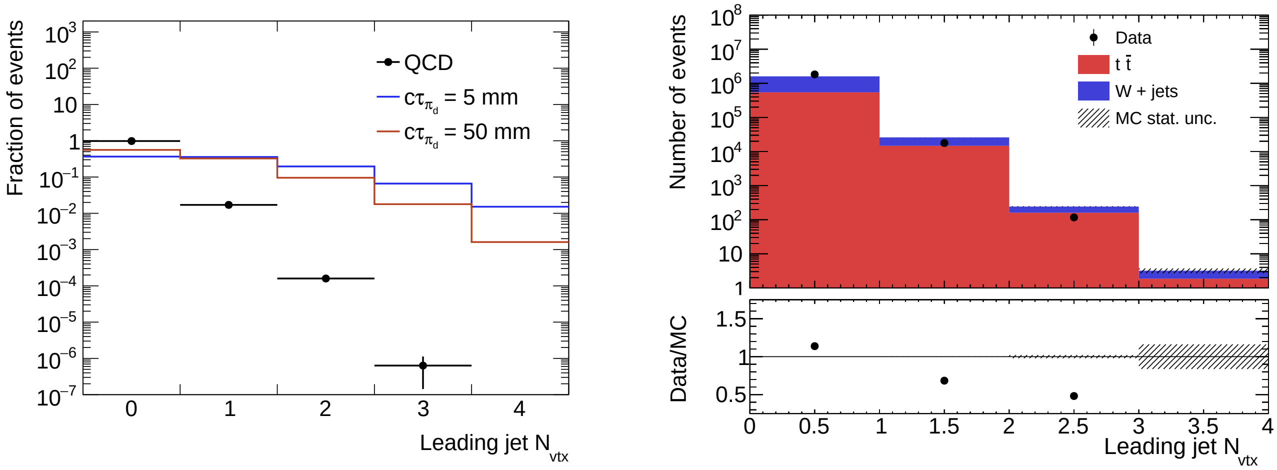
	Pre-selections
N_{jet} ($p_T > 200$ GeV, $\lvert\eta\rvert < 1.5$, overlap removal with photons)	≥ 2
m_{jj}	≤ 1 TeV
Trigger-matched jet	$p_T > 250$ GeV, PTF < 0.04
Leading and sub-leading jet ΔPTF	< 0.4

Table 5.4: Pre-selections applied in the low-m_{jj} channel of the cut-based approach.

5.6.2 Event selection

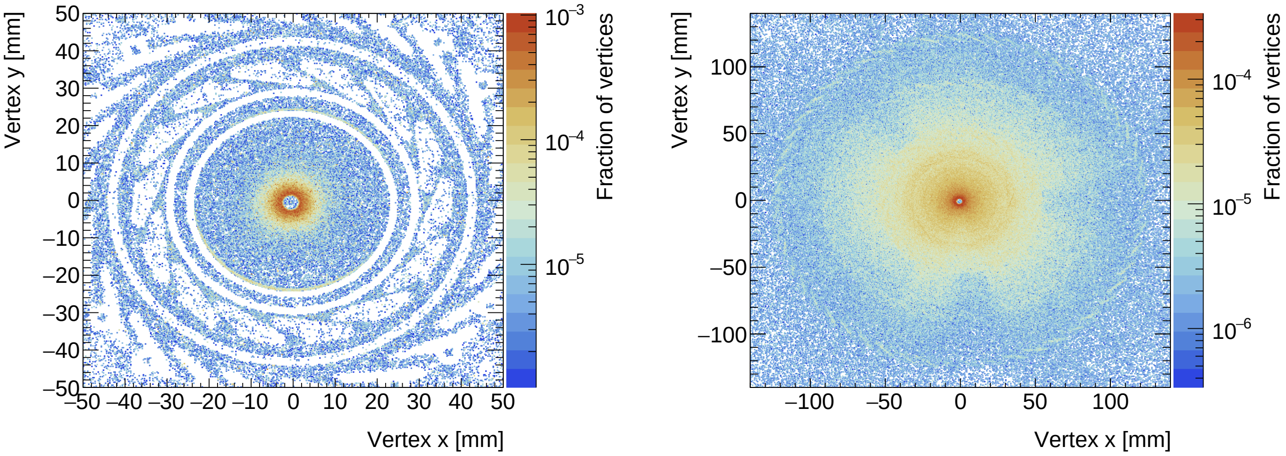
Given the nature of the emerging jets, two main types of jet observables can be exploited to discriminate them from background jets : track-based observables such as the PTF that rely on the fact that the emerging jets are likely to contain an important amount of large radius tracks, and jet-substructure observables that are related to the jet internal energy repartition. Indeed, as described in section 2.4, emerging jets arise from a double hadronization process, and thus have a highly different substructure from SM jets.

Moreover, the number of selected secondary vertices N_{vtx} as described in section 5.3.2 is also extremely discriminant between signal and background as shown in Figure 5.15.a. The distributions for data and MC events in the muon control region is shown in Figure 5.15.b, and present some disparities. The reason is that the veto applied to reject vertices in the detector material is deduced from an estimation of the detector position using data. On the other hand, the simulation of the detector presents small discrepancies with the real detector concerning the material position. In this way, it is possible that simulated displaced vertices located in simulated material areas are not vetoed. It is illustrated in Figure 5.16.a that presents the repartition of the secondary vertices from QCD events in the transverse plane of the detector : it is possible to notice some areas at a radius $\sim$ 25 mm corresponding to the simulated material that present an over density of vertices even after the veto being applied. However, it is not problematic for the analysis since the background estimation is done using data driven techniques, and also because the vertices from the signal are not correlated to the material, as illustrated in Figure 5.16.b that presents the repartition of the secondary vertices from signal events in the transverse plane. A selection $N_{\text{vtx}} \geq 1$ applied to both leading and sub-leading jets is sufficient to reduce considerably the number of MC background events while selecting a large fraction of signal events.



(a) (b)

Figure 5.15: N_{vtx} distributions (a) for QCD and signal events corresponding to $m_{Z'} = 0.6$ TeV and $m_{\pi_d} = 10$ GeV after the application of the emerging jet trigger and (b) for data and MC events in the muon control region.



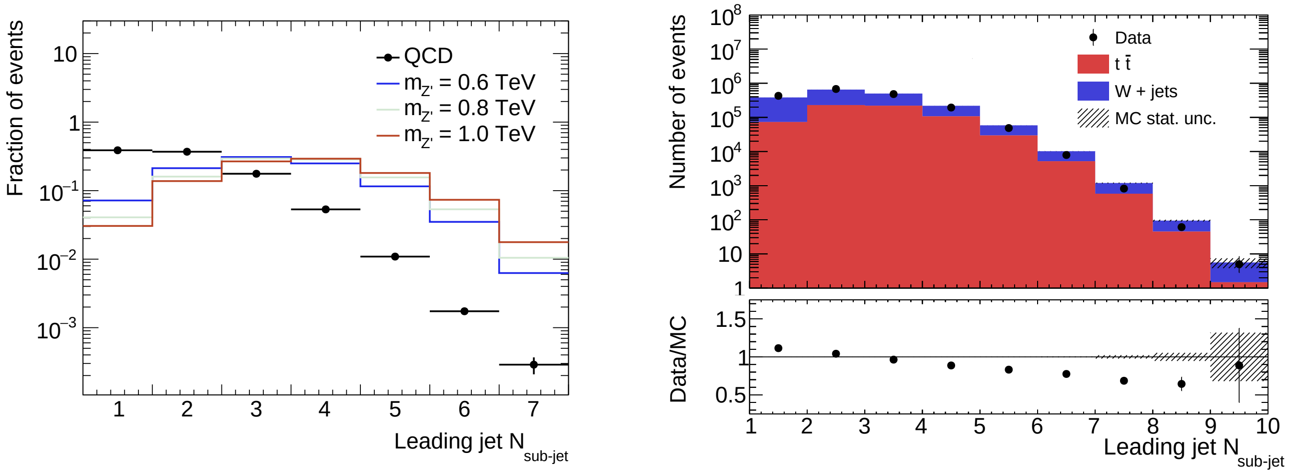
(a) (b)

Figure 5.16: Normalized distributions of the reconstructed secondary vertices in the transverse plane (x, y) from (a) QCD background events after the material map veto being applied and (b) signal events corresponding to $m_{Z'} = 1.5$ TeV, $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 50$ mm.

Given that a selection on the PTF is already applied to the trigger-matched jet in the preselections, a selection on another track-based variable quantifying the proportion of prompt (or displaced tracks) associated to the selected jets could be highly correlated. Instead, a selection on an observable related to the jet substructure will be complementary to the selections already applied before.

Several jet substructure variables computed from the associated tracks have been considered to discriminate the signal from background, such as ECF2 or Q_w (cf section 4.5.6). However, the variable that demonstrates the best discriminating power is the number of small-R sub-jets $N_{\text{sub-jet}}$ that served as constituents during the reclustering procedure to form large-R jet as described in section 5.3.1. Indeed, each dark pion contained in the emerging jets decays back to SM quarks that then produce SM jets. The internal structure of the emerging jets is thus supposed to consists in several sub-jets contained inside a larger jet. The distributions in background and signal events of $N_{\text{sub-jet}}$ for the leading jet is shown in Figure 5.17.a, in addition to the data and MC distributions in the muon control region in Figure 5.17.b. The agreement is globally correct in the bulk of the distribution between data and MC events.

The corresponding selection that defines the SR is $N_{\text{sub-jet}} \geq 3$ applied to both leading and sub-leading jets. The selections defining the SR are summarized in Table 5.5 : in a complementary way, they allow to select events by fully exploiting the particularities of the emerging jets. As it will be explained in the following section, the events passing all the selections except the one on $N_{\text{sub-jet}}$ will be used for background estimation.



(a) (b)

Figure 5.17: $N_{\text{sub-jet}}$ distributions (a) for QCD and signal events corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm after the application of the emerging jet trigger and (b) for data and MC events in the muon control region.

	Selections
Leading and sub-leading jet N_{vtx}	≥ 1
Leading and sub-leading jet $N_{\text{sub-jet}}$	≥ 3

Table 5.5: Selections applied in the low-m_{jj} channel of the cut-based approach.

The cutflows for MC background and signal events for two different $c\tau_{\pi_d}$ are presented in Table 5.6 and 5.7 respectively. The MC background events are normalized to an integrated luminosity of 51.8 fb^{-1} corresponding to the amount of data that is used in this analysis, while the numbers of signal events are unweighted and correspond to the numbers of generated events.

The cutflows for signal events present two notable differences between $c\tau_{\pi_d} = 5$ mm and $c\tau_{\pi_d} = 50$ mm considering the selection efficiencies. The selection on PTF has a better efficiency for the signal corresponding to $c\tau_{\pi_d} = 50$ mm, while the one on N_{vtx} select more efficially the signal corresponding to $c\tau_{\pi_d} = 5$ mm. Indeed, for two signal corresponding to the same $m_{Z'}$ and m_{π_d}, the one with the lowest value of $c\tau_{\pi_d}$ will produce more prompt tracks and thus lead to jets with higher PTF. On the other hand, as explained in section 4.2.2, the vertex reconstruction efficiency is reduced at large radius in the detector, leading to a lower efficiency for the selection on N_{vtx} for the signal corresponding to the higher dark pion lifetime.

	QCD di-jet	$t\bar{t}$	**Background**	Rejection [%]
Emerging jet trigger & $N_{\text{jet}} \geq 2$	19418139	35776	19453915	
Trigger-matched jet : $p_T > 250$ GeV & PTF < 0.04	4619505	7223	4626728	76
Leading & and sub-leading jets : $\lvert\eta\rvert < 1.5$, OR with γ & ΔPTF < 0.4	2321488	4271	2325759	50
Leading & and sub-leading jets : $N_{\text{vtx}} \geq 1$	545	3	548	100
$m_{jj} \leq 1$ TeV	470	2	472	14
Leading & and sub-leading jets : $N_{\text{sub-jet}} \geq 3$	2	1	3	99

Table 5.6: Cutflow for MC background events normalized to 51.8 fb^{-1}. The background rejection quantifies the fraction of background events that are rejected by a selection.

	$c\tau_{\pi_d} = 5$ mm	ϵ [%]	$c\tau_{\pi_d} = 50$ mm	ϵ [%]
Initial	200000		200000	
Emerging jet trigger & $N_{\text{jet}} \geq 2$	44585	22	50279	25
Trigger-matched jet : $p_T > 250$ GeV & PTF < 0.04	20843	47	33958	68
Leading & and sub-leading jets : $\lvert\eta\rvert < 1.5$, OR with γ & ΔPTF < 0.4	14818	71	25032	74
Leading & and sub-leading jets : $N_{\text{vtx}} \geq 1$	3820	26	2470	10
$m_{jj} \leq 1$ TeV	3754	98	2424	98
Leading & and sub-leading jets : $N_{\text{sub-jet}} \geq 3$	1845	49	1229	51

Table 5.7: Cutflow for signal events corresponding to $m_{Z'} = 600$ GeV, $m_{\pi_d} = 5$ GeV and $c\tau_{\pi_d} = 5, 50$ mm. The initial line corresponds to the number of generated events, and the efficiency of each selection is represented by ϵ.

5.7 Background estimation

5.7.1 ABCD method

To estimate the background contribution in the SR, a data driven method (i.e relying on data) named «ABCD method» is used. Indeed, given the very particular selections applied in this analysis, the background events selected in the SR are subject to MC simulation mis-modelling and are thus not reliable to estimate the background precisely.

The ABCD method relies on the definition of three distinct control regions (CRs) that are all orthogonal to the SR. More precisely, if the SR is defined with two selections, the CRs are obtained by inverting one or two of the selections. In this way, these CRs are enriched in background events while the signal is mainly populating the SR.

For example, if the SR, that is named region A, is defined with two continuous observables v_1 and v_2 such as $v_1 > 60$ and $v_2 > 50$, the three CRs named B, C and D can be defined as :

- Region B : $v_1 < 60$ and $v_2 > 50$,
- Region C : $v_1 > 60$ and $v_2 < 50$,
- Region D : $v_1 < 60$ and $v_2 < 50$.

The four regions form an «ABCD plane» which is schematized in Figure 5.18 [80], with data events in the three control regions in addition to the signal repartition in the plane. To predict the background contribution in the region A, the ABCD method relies on the assumption that v_1 and v_2 are independent for background events, leading to the formula :

$$N_{\mathrm{A}}^{\mathrm{bkg}} = N_{\mathrm{B}}^{\mathrm{bkg}} \times N_{\mathrm{C}}^{\mathrm{bkg}} / N_{\mathrm{D}}^{\mathrm{bkg}} \tag{5.4}$$

with $N_{\mathrm{X}}^{\mathrm{bkg}}$ the number of background events in the region X. Moreover, if the signal presence is negligible in these CRs, $N_{\mathrm{X}}^{\mathrm{bkg}}$ can be approximated by $N_{\mathrm{X}}^{\mathrm{obs}}$, the number of observed data events in the region X, leading to :

$$N_{\mathrm{A}}^{\mathrm{bkg}} \approx N_{\mathrm{B}}^{\mathrm{obs}} \times N_{\mathrm{C}}^{\mathrm{obs}} / N_{\mathrm{D}}^{\mathrm{obs}}. \tag{5.5}$$

With this method, the background contribution in the SR can be estimated by only considering the data in the CRs with the SR said to be «blinded».

Moreover, instead of considering that $N_{\mathrm{X}}^{\mathrm{bkg}} \approx N_{\mathrm{X}}^{\mathrm{obs}}$ in the CRs, a simultaneous likelihood fit can be performed to adjust the value of $N_{\mathrm{X}}^{\mathrm{bkg}}$ to the data by taking into account the signal event repartition in the ABCD plane. This will be detailed in section 5.9.

The ABCD method is used to estimate background for the cut-based approach in both the low and high-m_{jj} channels.

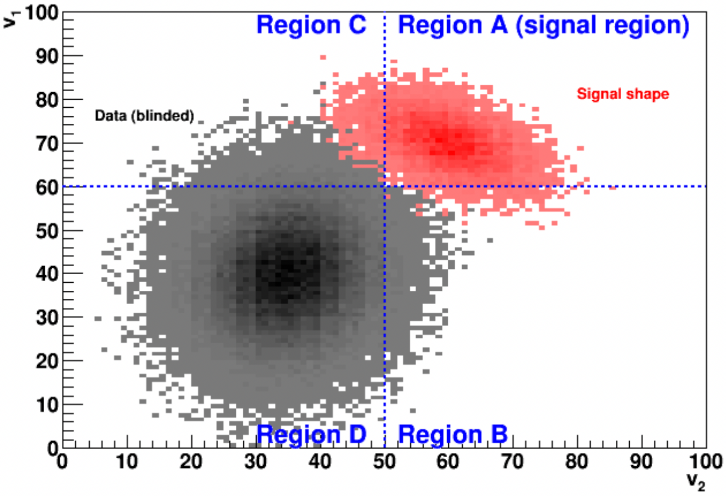


Figure 5.18: Illustrative ABCD plane. The black distribution corresponds to blinded data events (i.e not including the events in region A), and the red one correspond to signal events [80].

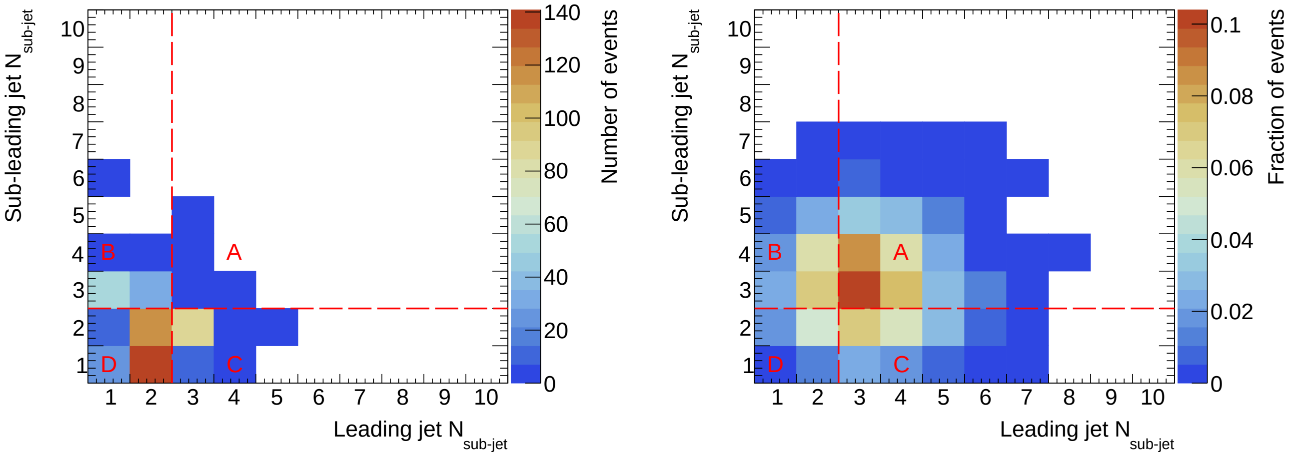
In the low-m_{jj} channel, the variables used to defined the ABCD plane are the number of sub-jets for the leading and the sub-leading jet $N^{\text{lead.}}_{\text{sub-jet}}$ and $N^{\text{sub-lead.}}_{\text{sub-jet}}$. Indeed, regarding Tables 5.6 and 5.7, $N_{\text{sub-jet}}$ demonstrates a strong discriminating power with a good balance between background rejection and signal selection. Moreover, choosing a variable computed from different jets can help to minimize direct correlation.

The four regions are defined as :

- Region A : $N^{\text{lead.}}_{\text{sub-jet}} \geq 3$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} \geq 3$,
- Region B : $N^{\text{lead.}}_{\text{sub-jet}} < 3$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} \geq 3$,
- Region C : $N^{\text{lead.}}_{\text{sub-jet}} \geq 3$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} < 3$,
- Region D : $N^{\text{lead.}}_{\text{sub-jet}} < 3$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} < 3$.

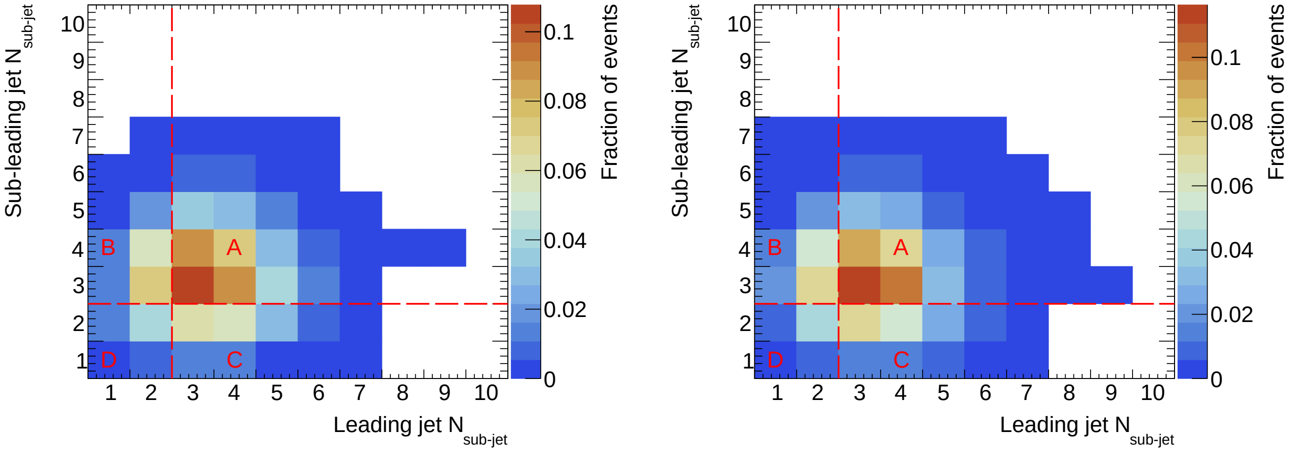
These regions are filled with events passing all the selections detailed previously except the one on $N_{\text{sub-jet}}$. The distributions in the ABCD plane of the MC background and signal events are illustrated in Figures 5.19, 5.20 and 5.21, and show a satisfying separation between background and signal events.

The MC background event yields in the four regions is presented in Table 5.8. It can be observed that the statistical uncertainties on the different numbers of events are large : it is due both to the limited number of events available and to several MC events having large weights. It is thus difficult to conclude on the independence of the two variables based on MC events. As a consequence, the validation of the ABCD method, i.e the verification of the validity of the equation (5.4), has been done in data in a signal free ABCD plane as it will be detailed in the following.



(a) (b)

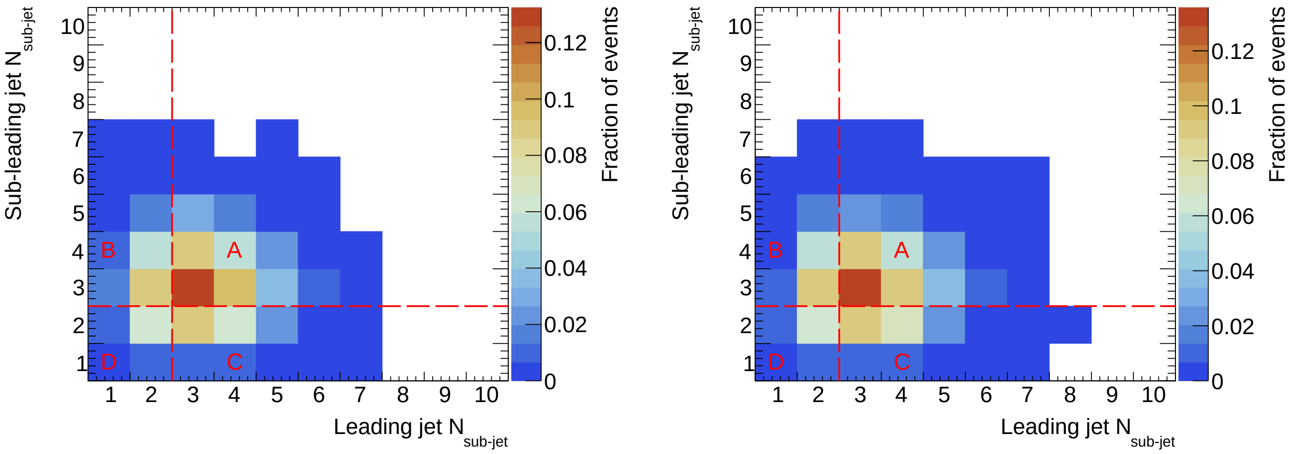
Figure 5.19: ABCD plane (a) for MC background and (b) signal events corresponding to $m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 5$ mm and $m_{Z'} = 600$ GeV.



(a) (b)

Figure 5.20: ABCD plane for signal events corresponding to (a) $m_{\pi_d} = 10$ GeV, $c\tau_{\pi_d} = 5$ mm and $m_{Z'} = 600$ GeV and (b) $m_{\pi_d} = 10$ GeV, $c\tau_{\pi_d} = 50$ mm and $m_{Z'} = 600$ GeV.

The repartition of the signal events in the four regions is presented in Table 5.9. For each signal, it corresponds to $N_{\mathrm{X}}/(N_{\mathrm{A}} + N_{\mathrm{B}} + N_{\mathrm{C}} + N_{\mathrm{D}})$, where N_{X} is the number of signal events in the region $\mathrm{X} \in \{\mathrm{A, B, C, D}\}$. As expected, the signal events are mainly populating the region A.



(a) (b)

Figure 5.21: ABCD plane for signal events corresponding to (a) $m_{\pi_d} = 20$ GeV, $c\tau_{\pi_d} = 5$ mm and $m_{Z'} = 600$ GeV and (b) $m_{\pi_d} = 20$ GeV, $c\tau_{\pi_d} = 50$ mm and $m_{Z'} = 600$ GeV.

	QCD di-jet	$t\bar{t}$	**Background**
A	2.1 ± 1.5	0.6 ± 0.2	2.7 ± 1.5
B	84 ± 38	0.6 ± 0.2	85 ± 38
C	98 ± 41	0.7 ± 0.2	99 ± 41
D	280 ± 110	0.6 ± 0.2	280 ± 110
B × C/D			30 ± 22

Table 5.8: MC background event yields in the ABCD plane with the associated statistical uncertainties.

	A	B	C	D
$m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 5$ mm	49	21	21	9
$m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 50$ mm	51	20	21	8
$m_{\pi_d} = 10$ GeV, $c\tau_{\pi_d} = 5$ mm	56	19	19	6
$m_{\pi_d} = 10$ GeV, $c\tau_{\pi_d} = 50$ mm	56	19	19	6
$m_{\pi_d} = 20$ GeV, $c\tau_{\pi_d} = 5$ mm	51	20	20	9
$m_{\pi_d} = 20$ GeV, $c\tau_{\pi_d} = 50$ mm	51	19	22	8

Table 5.9: Repartition in the ABCD plane of the signal events corresponding to $m_{Z'} = 600$ GeV. The numbers are in %.

5.7.2 Validation

To validate the ABCD method, an alternative ABCD plane is used. It exploits the jet classification score provided by the tagging algorithm that is used in the ML-based approach. Indeed, this tagging algorithm has been trained to be able to differentiate QCD jets from emerging ones. It has been developed to output, for a given jet, a classification score corresponding to the probability that the considered jet is an emerging one. The distribution of this jet classification score for data, MC background and signal events is shown in Figure 5.22 [57].

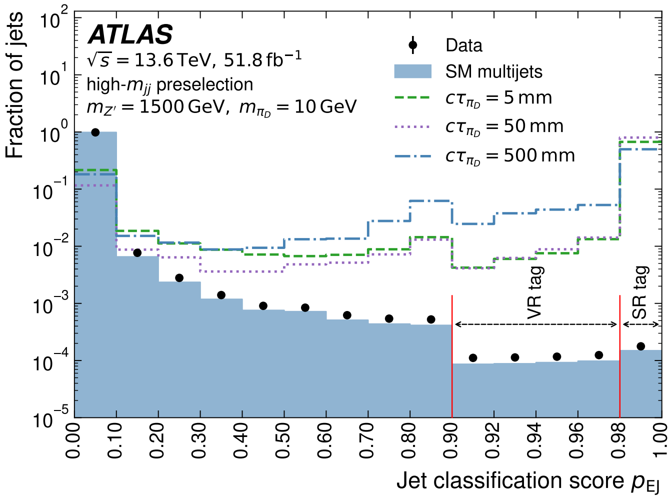


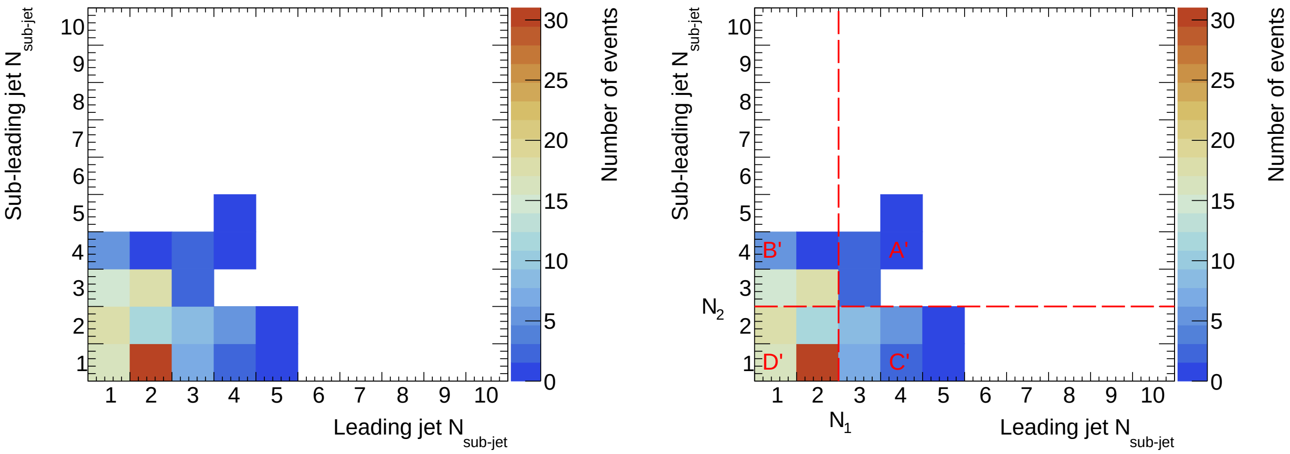
Figure 5.22: Normalized distributions of the jet classification score for jets from data, simulated multijets and signal events passing the ML-based high-m_{jj} channel pre-selections [57].

A signal-free ABCD plane can thus be defined by applying the same selections as the nominal ABCD plane and requiring that both leading and sub-leading jets have a score below 0.95. The MC background event yields in this signal-free ABCD plane are nearly unchanged with respect to the yields in the nominal one as shown in Table 5.10, while the signal events are completely removed.

Signal-free	QCD di-jet	$t\bar{t}$	**Background**
A	2.1 ± 1.5	0.5 ± 0.2	2.6 ± 1.5
B	84 ± 38	0.4 ± 0.2	85 ± 38
C	98 ± 41	0.6 ± 0.2	99 ± 41
D	230 ± 110	0.5 ± 0.2	230 ± 110

Table 5.10: MC background event yields in the signal-free ABCD plane.

The repartition of the data events in the signal-free ABCD plane is shown in Figure 5.23.a, and the number of observed events in the four regions defined with the nominal selections on $N_{\text{sub-jet}}$ are presented in Table 5.11.



(a)

(b)

Figure 5.23: Distribution of data events in the signal-free ABCD plane.

Signal-free	**Data**
A	8
B	39
C	25
D	77
B × C/D	12.7 ± 4.2 (stat.)

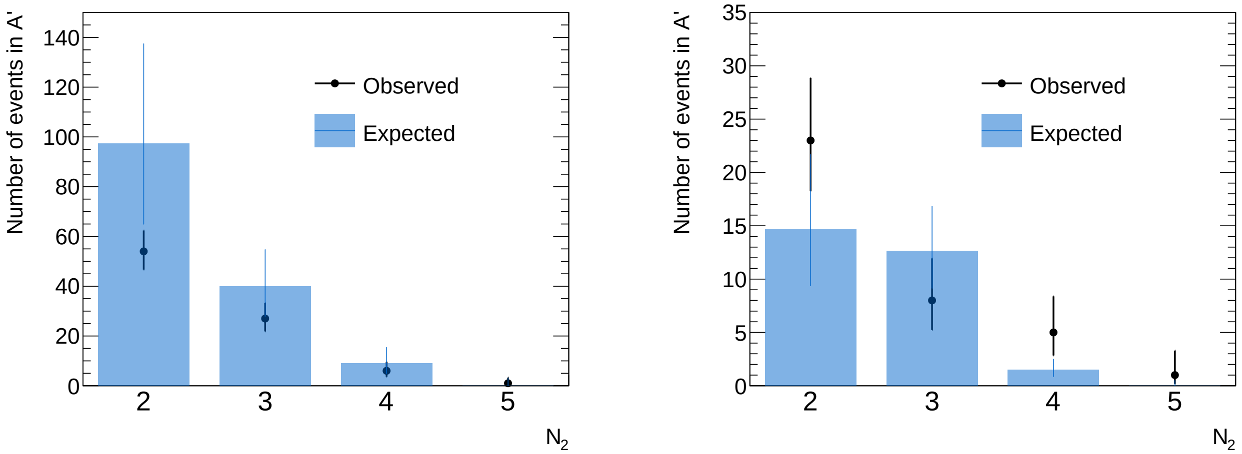
Table 5.11: Observed yields in the signal-free ABCD plane.

To validate the ABCD method, the principle is the following : four regions can be defined with two values N_1 and N_2 such as :

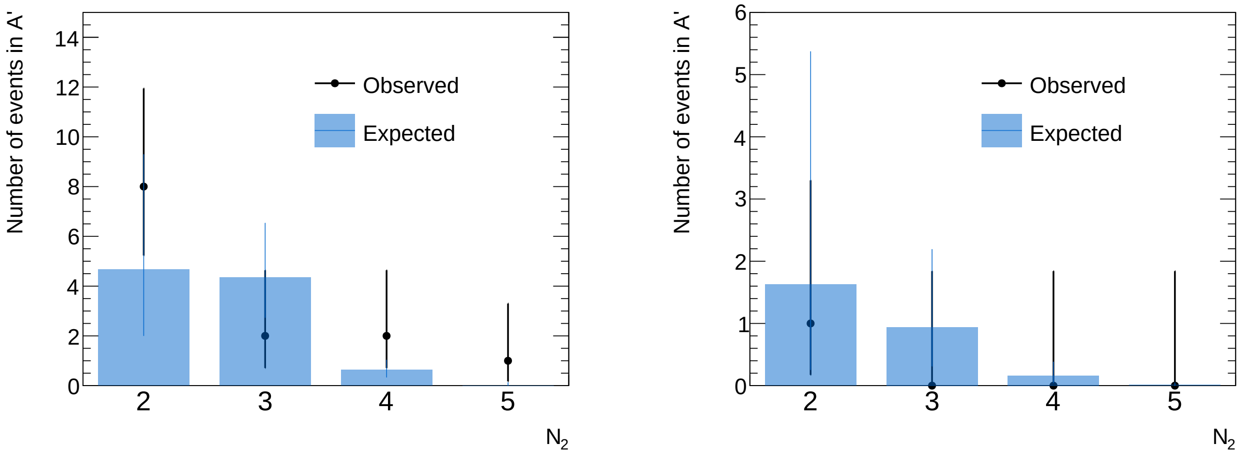
- Region A' : $N^{\text{lead.}}_{\text{sub-jet}} \geq N_1$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} \geq N_2$
- Region B' : $N^{\text{lead.}}_{\text{sub-jet}} < N_1$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} \geq N_2$
- Region C' : $N^{\text{lead.}}_{\text{sub-jet}} \geq N_1$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} < N_2$
- Region D' : $N^{\text{lead.}}_{\text{sub-jet}} < N_1$ and $N^{\text{sub-lead.}}_{\text{sub-jet}} < N_2$

as illustrated in Figure 5.23.b. Then, for a given couple (N_1, N_2), the number of observed events in A' can be compared to the expected number of events computed as $N^{\text{exp}}_{\text{A'}} = N_{\text{B'}} \times N_{\text{C'}}/N_{\text{D'}}$, with $N_{\text{X'}}$ the observed number of events in X'. This comparison has been done with all the possible values for N_1 and N_2 as shown in Figure 5.24.

The available statistics is extremely low and leads to important statistical uncertainties, but the predicted numbers of events remain compatible with the observed ones. However, a non-closure systematic uncertainty on the prediction is derived as described in the next section.



(a) $N_1 = 2$ (b) $N_1 = 3$



(c) $N_1 = 4$ (d) $N_1 = 5$

Figure 5.24: Comparison between the observed and expected number of events in A' with all the possible values for (N_1, N_2) being scanned.

5.7.3 Non-closure uncertainty

This non-closure uncertainty is determined using data in the nominal ABCD plane, i.e with the selection on the jet classification score removed in order to determine this uncertainty independently from the ML-based approach.

Setting a non-closure uncertainty may also help to take into account a potential indirect dependence between the number of sub-jets for the leading and sub-leading jets. Indeed, given that $N_{\text{sub-jet}}$ is potentially correlated to event observables such as μ (cf section 3.3.2) or the number of primary vertex, $N_{\text{sub-jet}}^{\text{lead.}}$ and $N_{\text{sub-jet}}^{\text{sub-lead.}}$ may be indirectly dependent in data.

The uncertainty has been derived as $f \times \sigma_{\text{stat}}(N_{\text{A}}^{\text{exp}})$ with $\sigma_{\text{stat}}(N_{\text{A}}^{\text{exp}})$ the statistical uncertainty in the expected background in the SR. To determine f, data events in regions B, C and D have been considered.

First in B and D regions, i.e with events passing $N_{\text{sub-jet}}^{\text{lead.}} < 3$, four regions are defined such as :

- Region R_1 : $N_{\text{sub-jet}}^{\text{lead.}} = 2$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} \geq N$
- Region R_2 : $N_{\text{sub-jet}}^{\text{lead.}} = 1$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} \geq N$
- Region R_3 : $N_{\text{sub-jet}}^{\text{lead.}} = 2$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} < N$
- Region R_4 : $N_{\text{sub-jet}}^{\text{lead.}} = 1$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} < N$

with N varying from 2 to 5 as illustrated in Figure 5.25. For each N, the prediction in R_1 is computed such as $N_{R_1}^{\text{exp}} = N_{R_2} \times N_{R_3}/N_{R_4}$ with N_{R_X} the number of observed events in R_X. Moreover, the quantity $|N_{R_1}^{\text{exp}} - N_{R_1}|/\sqrt{\sigma_{\text{stat}}^2(N_{R_1}^{\text{exp}}) + \sigma_{\text{stat}}^2(N_{R_1})}$, with $\sigma_{\text{stat}}(N_{R_1}^{\text{exp}})$ the statistical uncertainty on $N_{R_1}^{\text{exp}}$ and $\sigma_{\text{stat}}(N_{R_1})$ the one on N_{R_1}, is computed (cf Table 5.12). It represents the significance of the difference between the observed and expected yields in R_1.

Moreover in C and D with events passing $N_{\text{sub-jet}}^{\text{sub-lead.}} < 3$, four regions are defined such as :

- Region $R_{1'}$: $N_{\text{sub-jet}}^{\text{lead.}} \geq N'$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} = 2$
- Region $R_{2'}$: $N_{\text{sub-jet}}^{\text{lead.}} < N'$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} = 2$
- Region $R_{3'}$: $N_{\text{sub-jet}}^{\text{lead.}} \geq N'$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} = 1$
- Region $R_{4'}$: $N_{\text{sub-jet}}^{\text{lead.}} < N'$ and $N_{\text{sub-jet}}^{\text{sub-lead.}} = 1$

with N' varying from 2 to 5 as illustrated in Figure 5.26. Similarly, for each N', the prediction in $R_{1'}$ given by $N_{R_{1'}}^{\text{exp}} = N_{R_{2'}} \times N_{R_{3'}}/N_{R_{4'}}$ and $|N_{R_{1'}}^{\text{exp}} - N_{R_{1'}}|/\sqrt{\sigma_{\text{stat}}^2(N_{R_{1'}}^{\text{exp}}) + \sigma_{\text{stat}}^2(N_{R_{1'}})}$ are computed (cf Table 5.13).

Finally, f is computed as the average of the quantity $|N^{\text{exp}} - N|/\sqrt{\sigma_{\text{stat}}^2(N^{\text{exp}}) + \sigma_{\text{stat}}^2(N)}$ across all the R_1 and $R_{1'}$ regions, giving an f value of around 1. The non-closure uncertainty thus corresponds to $1.0 \times \sigma_{\text{stat}}(N_{\text{A}}^{\text{exp}})$, and is taken into account as a systematic uncertainty during the statistical interpretation (cf section 5.9).

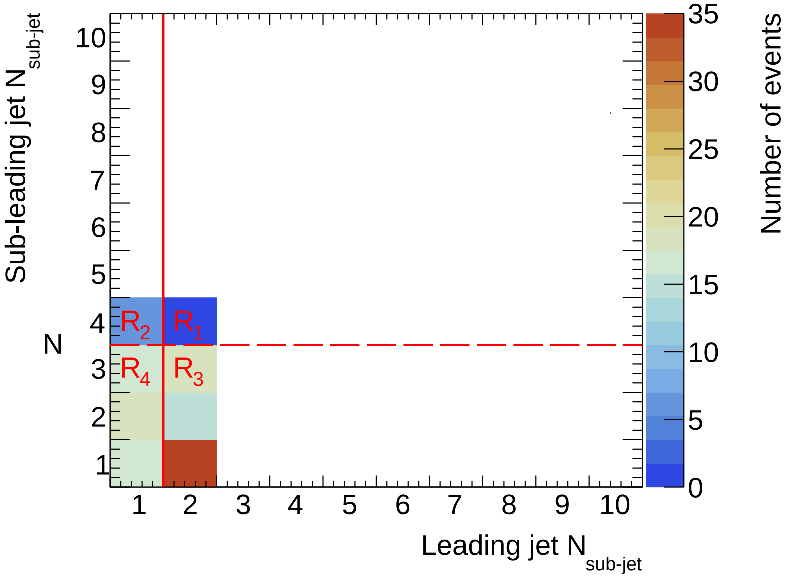


Figure 5.25: Regions B and D used to derive the non-closure uncertainty.

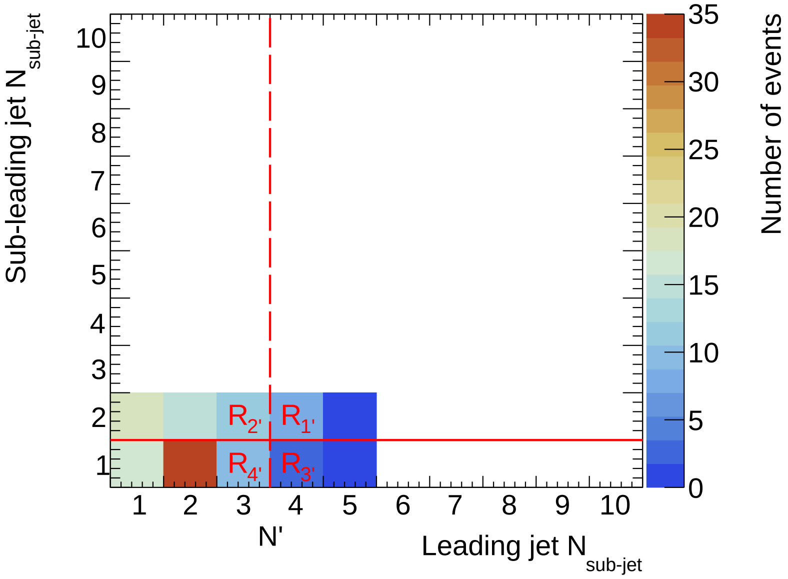


Figure 5.26: Regions C and D used to derive the non-closure uncertainty.

B and D	N_{R_1}	$N_{R_1}^{\text{exp}}$	$\sigma_{\text{stat}}(N_{R_1})$	$\sigma_{\text{stat}}(N_{R_1}^{\text{exp}})$	$\lvert N_{R_1} - N_{R_1}^{\text{exp}}\rvert / \sqrt{\sigma_{\text{stat}}^2(N_{R_1}) + \sigma_{\text{stat}}^2(N_{R_1}^{\text{exp}})}$ [%]
$N = 2$	34	84.4	6.9	34.4	143.5
$N = 3$	20	32.2	5.5	11.7	94.3
$N = 4$	1	7.8	2.3	5.0	125.1
$N = 5$	0	0.0	1.8	2.2	0.0
Average					**90.7**

Table 5.12: Number of observed and expected events in addition to the associated statistical uncertainties in each possible R_1 region used to derive the non-closure uncertainty.

C and D	$N_{R_{1'}}$	$N_{R_{1'}}^{\text{exp}}$	$\sigma_{\text{stat}}(N_{R_{1'}})$	$\sigma_{\text{stat}}(N_{R_{1'}}^{\text{exp}})$	$\lvert N_{R_{1'}} - N_{R_{1'}}^{\text{exp}}\rvert / \sqrt{\sigma_{\text{stat}}^2(N_{R_{1'}}) + \sigma_{\text{stat}}^2(N_{R_{1'}}^{\text{exp}})}$ [%]
$N' = 2$	35	50.8	7.0	23.2	65.3
$N' = 3$	21	8.0	5.7	3.6	194.3
$N' = 4$	9	2.9	4.1	2.4	128.8
$N' = 5$	1	0.8	2.3	1.9	6.3
Average					**98.7**

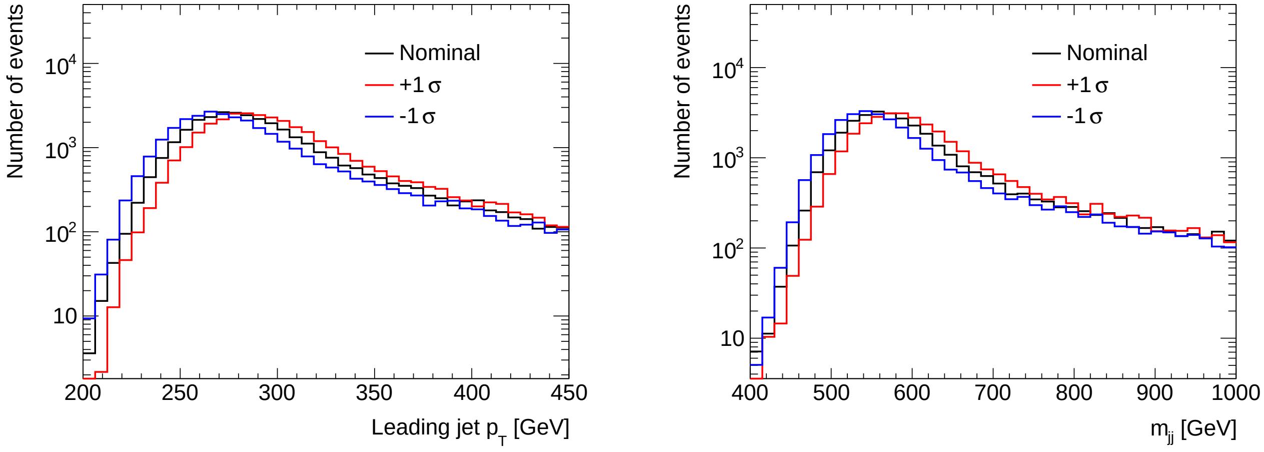
Table 5.13: Number of observed and expected events in addition to the associated statistical uncertainties in each possible $R_{1'}$ region used to derive the non-closure uncertainty.

5.8 Signal systematic uncertainties

The systematic uncertainties concerning the modelling of the signal events have to be considered. They correspond either to experimental uncertainties associated to the reconstruction of the physical objects that are used, or to theoretical uncertainties related to the MC particle generator. Each of these systematic uncertainties affects the signal acceptance.

5.8.1 Experimental

As explained in section 5.3.1, it is the small-R jets that are calibrated. The systematic uncertainties related to the calibration of the small-R jets are thus propagated to the large-R jets. They consists either in uncertainty on the jet energy scale (JES) or on the jet energy resolution (JER) that arise from differences between data and MC. For each systematics related to the JES, the jet energy nominal values are varied according to positive and negative variations as illustrated in Figure 5.27. On the other hand, for each systematics related to the JER, a smearing is applied to the jet energy response distributions.



(a) (b)

Figure 5.27: Impact of a JES systematic uncertainty on (a) the jet p_T and on (b) the dijet invariant mass m_{jj} for signal events corresponding to $m_{Z'} = 0.6$ TeV, $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm.

The systematics related to the tracks allow to take into account the uncertainty on the track reconstruction efficiency or on the evaluation of the track impact parameters d_0 and z_0. For each of these systematic uncertainties, a collection of tracks varied from the nominal one is obtained and used to re-compute track-based observables like the PTF.

The track systematic uncertainties also impact the vertex reconstruction. By combining all the track systematics, a collection of secondary vertices varied from the nominal one is obtained, and N_{vtx} can be re-computed.

Moreover, the application of the weight $w_{\text{pile-up}}$ in the equation (5.1), procedure that is named pile-up re-weighting (PRW), has a corresponding systematic that needs to be considered. Indeed, the application of $w_{\text{pile-up}}$ aims to obtain the same distribution of μ between data and MC events. However, given that differences exist in the evaluation of μ between data and MC, the PRW procedure is affected by a systematic uncertainty. This one is taking into account through up and down variations of $w_{\text{pile-up}}$, leading to modified MC event weights.

Finally, an uncertainty in the total integrated luminosity is also considered, which is about 2% for the data collected in 2022 and 2023.

5.8.2 Theoretical

The theoretical uncertainties are related to the uncertainties in the QCD parameters that are used during the MC simulations. These uncertainties are related to the renormalization and factorization scales and α_S value, and to the parton shower tuning parameters that determine the quantity of ISR and FSR in the events that impact the jet sub-structure. Each theoretical systematic is taken into account through varied MC event weights w_{MC} that are produced by the particle generator.

5.8.3 Total

For each systematic and for a given signal, the varied event yield in the SR is computed. It can be greater or lower than the nominal yield, and thus lead to positive or negative relative variation.

To each JES systematic corresponds a $+1\sigma$ and a -1σ variation. The positive (negative) relative variations to the nominal yield are summed in quadrature in order to compute the global up (down) variation coming from the jet systematic uncertainties. Similarly, there is a positive and a negative variation related to the single PRW uncertainty. Ultimately, for both jet and PRW systematics, the final relative variation is taken to be the greatest variation in absolute value between the positive and the negative one.

On the other hand for the track and JER systematics, only negative variations are considered. As a consequence, the down variations are symmetrized to obtain the up ones. Similarly as for the jet systematics, the relative variations to the nominal yield due to track uncertainties are summed in quadrature to obtain the total variation.

Concerning the theoretical uncertainties, the procedure is different. Indeed in this case, the different systematics correspond to variations of the same set of QCD parameters. As a consequence, the relative variations due to each theoretical systematic are not summed, but it is instead the greatest variation in absolute value that is symmetrized and taken to be the total theoretical uncertainty.

The Tables 5.14, 5.15 and 5.16 present for different signals the relative variations due to each type of systematics, in addition to the total uncertainty which is obtained by summing in quadrature all the variations.

	Theoretical	Jet	Track	Vertex	PRW	**Total**
$m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 5$ mm	± 5.0	± 21.9	± 1.0	± 4.2	± 0.8	± 22.9
$m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 50$ mm	± 7.1	± 15.3	± 0.3	± 6.9	± 3.7	± 18.6
$m_{\pi_d} = 10$ GeV, $c\tau_{\pi_d} = 5$ mm	± 6.4	± 22.0	± 1.8	± 3.0	± 0.3	± 23.2
$m_{\pi_d} = 10$ GeV, $c\tau_{\pi_d} = 50$ mm	± 6.2	± 13.8	± 0.7	± 5.4	± 2.6	± 16.3
$m_{\pi_d} = 20$ GeV, $c\tau_{\pi_d} = 5$ mm	± 6.7	± 17.2	± 3.1	± 3.0	± 1.0	± 18.9
$m_{\pi_d} = 20$ GeV, $c\tau_{\pi_d} = 50$ mm	± 8.6	± 14.4	± 1.3	± 5.1	± 1.9	± 17.7

Table 5.14: Relative variation (%) on the signal yield in the SR for different signals corresponding to $m_{Z'} = 0.6$ TeV.

	Theoretical	Jet	Track	Vertex	PRW	**Total**
$m_{Z'} = 0.6$ GeV	± 6.4	± 22.0	± 1.8	± 3.0	± 0.3	± 23.2
$m_{Z'} = 0.8$ GeV	± 5.6	± 7.4	± 2.3	± 2.8	± 1.6	± 10.1
$m_{Z'} = 1.0$ GeV	± 5.2	± 3.7	± 1.6	± 2.6	± 1.7	± 7.3

Table 5.15: Relative variation (%) on the signal yield in the SR for different signals corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm.

	Theoretical	Jet	Track	Vertex	PRW	**Total**
$c\tau_{\pi_d} = 1$ mm	± 6.6	± 31.0	± 6.1	± 4.9	± 0.5	± 32.6
$c\tau_{\pi_d} = 5$ mm	± 6.4	± 22.0	± 1.8	± 3.0	± 0.3	± 23.2
$c\tau_{\pi_d} = 10$ mm	± 7.0	± 18.8	± 2.7	± 4.4	± 2.7	± 20.9
$c\tau_{\pi_d} = 50$ mm	± 6.2	± 13.8	± 0.7	± 5.4	± 2.6	± 16.3
$c\tau_{\pi_d} = 100$ mm	± 9.2	± 17.0	± 2.8	± 6.0	± 4.1	± 20.8

Table 5.16: Relative variation (%) on the signal yield in the SR for different signals corresponding to $m_{Z'} = 0.6$ TeV and $m_{\pi_d} = 10$ GeV.

It can be noticed that the jet systematics are largely dominant for signal corresponding to $m_{Z'} = 0.6$ TeV : indeed, systematic variations of the jet p_T distribution can lead to important variations in terms of acceptance since for $m_{Z'} = 0.6$ TeV signal, the peak of the distribution is close to the 250 GeV p_T threshold that is applied. It is also possible to see that for signals corresponding to $m_{Z'} = 0.8$ and 1.0 TeV, the jet systematics have a reduced impact since the peaks of the corresponding jet p_T distributions are situated at higher values. It is also possible to see that the track systematics are more impacting signals corresponding to lower $c\tau_{\pi_d}$, since the produced jets are expected to contain a larger amount of tracks than the ones coming from higher $c\tau_{\pi_d}$ signals.

For each signal, the total relative variation is taken as a systematic uncertainty during the statistical interpretation.

5.9 Statistical interpretation

The numbers of observed events in the four regions of the ABCD plane are presented in Table 5.17, in addition to the background prediction in the SR obtained with the application of the equation (5.5). It can be seen that the observed number of events in the SR is in agreement with the prediction within the total uncertainty, which corresponds to the sum in quadrature of the statistical and systematic uncertainties.

	Data
A	10
B	43
C	34
D	84
B × C/D	17.4 ± 5.1 (stat.) ± 5.1 (syst.) = 17.4 ± 7.2 (total)

Table 5.17: Observed yields in the ABCD plane.

As there are only 10 events in the SR, their properties are presented in Table 5.18.

p_T [GeV]	η	ϕ	PTF	$N_{\text{sub-jet}}$	m_{jj} [GeV]
(277.0, 202.4)	(-1.16, -0.61)	(2.16, -0.92)	(0.033, 0.412)	(3, 4)	526.4
(291.4, 248.3)	(0.43, 1.29)	(0.81, -2.42)	(0.02, 0.292)	(3, 3)	610.7
(425.5, 234.9)	(0.19, -1.4)	(-0.73, 2.17)	(0.039, 0.489)	(4, 4)	868.2
(414.4, 397.1)	(0.46, -0.16)	(-2.04, 1.31)	(0.027, 0.004)	(3, 3)	871.5
(298.0, 215.1)	(-0.33, 1.05)	(-0.72, 1.84)	(0.032, 0.719)	(3, 3)	631.0
(290.7, 287.6)	(-0.76, 0.48)	(0.37, -1.74)	(0.035, 0.495)	(4, 5)	687.4
(295.1, 255.5)	(-0.45, 1.45)	(1.55, -1.48)	(0.19, 0.009)	(3, 3)	838.1
(309.5, 270.9)	(1.35, 1.05)	(-2.54, 0.69)	(0.447, 0.02)	(3, 4)	620.4
(395.9, 377.5)	(-0.72, -0.02)	(2.97, -0.02)	(0.03, 0.629)	(3, 3)	873.2
(408.4, 373.6)	(0.21, -0.69)	(-1.86, 1.28)	(0.027, 0.364)	(3, 4)	908.6

Table 5.18: Events in the low-m_{jj} cut-based SR. For a jet variable X, the numbers are such that $(X^{\text{lead.}}, X^{\text{sub-lead.}})$, i.e the first (second) number in the parenthesis corresponds to the (sub-)leading jet.

These results need then to be interpreted in order to set upper limits on the production cross-section of the signal $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$, or to exclude values for the different parameters of the theoretical benchmark model. This interpretation is performed through a statistical test based on a likelihood function.

5.9.1 Statistical model

With $\mathcal{P}$ and $\mathcal{G}$ representing respectively Poisson and Gaussian distributions, the likelihood function used for the statistical interpretation is given by :

$$\mathcal{L}(\mu,\theta) = \prod_{\mathrm{X}\in\{\mathrm{A,\ B,\ C,\ D}\}} \mathcal{P}(N_{\mathrm{X}}|N_{\mathrm{X}}^{\mathrm{exp}}(\mu,\theta)) \times \prod_{i} \mathcal{G}(\theta_i) \tag{5.6}$$

where μ is the signal strength, θ represents nuisance parameters that encode the set of systematic uncertainties both on the signal and background prediction, and N_{X} and $N_{\mathrm{X}}^{\mathrm{exp}}$ represent respectively the number of observed and expected events in the region X. The expected yield $N_{\mathrm{X}}^{\mathrm{exp}}$ is equal to $\mu N_{\mathrm{X}}^{\mathrm{sig}} + N_{\mathrm{X}}^{\mathrm{bkg}}$ where $N_{\mathrm{X}}^{\mathrm{sig}}$ is the number of expected signal event and $N_{\mathrm{X}}^{\mathrm{bkg}}$ the expected background yield. $\mathcal{P}(N_{\mathrm{X}}|N_{\mathrm{X}}^{\mathrm{exp}})$ is thus representing the Poisson probability of observing N_{X} events given the expected yield $N_{\mathrm{X}}^{\mathrm{exp}}$, and $\mathcal{G}(\theta_i)$ corresponds to Gaussian constraint on the nuisance parameter θ_i.

The number of background events are expressed using three free parameters μ_B, τ_B and τ_C such as :

$$\begin{aligned} N_{\mathrm{A}}^{\mathrm{bkg}} &= \mu_B \\ N_{\mathrm{B}}^{\mathrm{bkg}} &= \mu_B \times \tau_B \\ N_{\mathrm{C}}^{\mathrm{bkg}} &= \mu_B \times \tau_C \\ N_{\mathrm{D}}^{\mathrm{bkg}} &= \mu_B \times \tau_B \times \tau_C \end{aligned}$$

and thus imposing the relation from the equation (5.4). In addition to μ, the likelihood function thus involves four free parameters.

The numbers of observed events N_{X} are provided to the likelihood function, in addition to the numbers of signal events $N_{\mathrm{X}}^{\mathrm{sig}}$ in the four regions. The likelihood fit consists then in finding the set of parameters μ, μ_B, τ_B and τ_C that maximized the function. In this way, the post-fit background prediction in the signal region $N_{\mathrm{A}}^{\mathrm{bkg}}$ can be different from the one presented in Table 5.17 that has been obtained by computing $N_{\mathrm{B}} \times N_{\mathrm{C}}/N_{\mathrm{D}}$.

5.9.2 Test statistic for upper limits

Given that no excess has been observed, a test statistic must be performed to find the maximum allowed value for the signal strength μ_{up} regarding the observed data. This upper limit is such that any value of μ higher than μ_{up} is excluded at a given confidence level, which is usually set to 95 %.

The definition of the test statistic is based on the profile likelihood ratio $\lambda(\mu)$ [81] that corresponds to :

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \tag{5.7}$$

with $\hat{\hat{\theta}}(\mu)$ corresponding to the value of θ that maximizes $\mathcal{L}$ for a given μ, and $\hat{\mu}$ and $\hat{\theta}$ are the parameters that maximize the likelihood function.

In this way, $\lambda(\mu)$ is a number that is comprised between 0 and 1, with λ close to 1 implying a great agreement between the data and the tested value of μ.

Then, for $\hat{\mu} > 0$, the test statistic for upper limits q_μ is defined as :

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{5.8}$$

In this way, higher values of q_μ correspond to important incompatibilities between the data and the tested value of μ.

The level of agreement between the data and a hypothesized μ is quantified with the p-value that is expressed as :

$$p_\mu = \int_{q_{\mu,\text{ obs}}}^{\infty} f(q_\mu|\mu)dq_\mu \tag{5.9}$$

with $q_{\mu,\text{ obs}}$ the value of the test statistic observed in data for a given μ, and $f(q_\mu|\mu)$ the probability density function of q_μ by assuming the hypothesis μ in data. In this way, a large value of p_μ corresponds to an important probability that the observation is compatible with μ.

Then, the CL_s [82] is defined as :

$$\text{CL}_s = \frac{p_\mu}{1 - p_b} \tag{5.10}$$

with $1 - p_b = \int_{q_{\mu,\text{ obs}}}^{\infty} f(q_\mu|0)dq_\mu$, and decreases with μ. The observed upper limit on the signal strength is finally taken as the maximum value of μ for which $\text{CL}_s < 1 - \alpha$, with α the confidence level (CL). In general, α is set to 95% and thus $1 - \alpha$ is equal to 0.05. This upper limit on the signal strength can then be translated into an upper limit on the signal cross-section.

In order to evaluate p_μ and p_b, it is necessary to know the distributions of $f(q_\mu|\mu')$, where μ corresponds to the value that is tested and μ' is the one that is assumed in data, especially with $\mu' = \mu$ and $\mu' = 0$. These distributions are obtained using the asymptotic formulae approximation as referred in [81].

Finally, expected upper limits and corresponding ±1 and $\pm 2\sigma$ bands are computed also with the CL_s by considering Asimov datasets, i.e datasets built by setting the observed yields equal to the expected ones.

5.9.3 Results

The construction of the statistical model and the implementation of the likelihood function has been made with the pyhf framework [83, 84].

The post-fit numbers of background events are presented in Table 5.19. These numbers has been obtained by considering signals corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm, but they are similar among the different signals. Given the extremely low post-fit value for $\mu \times N_\text{A}^\text{sig}$ that represents the number of signal events in the SR, N_X^exp is nearly equal to N_X^bkg in the four regions.

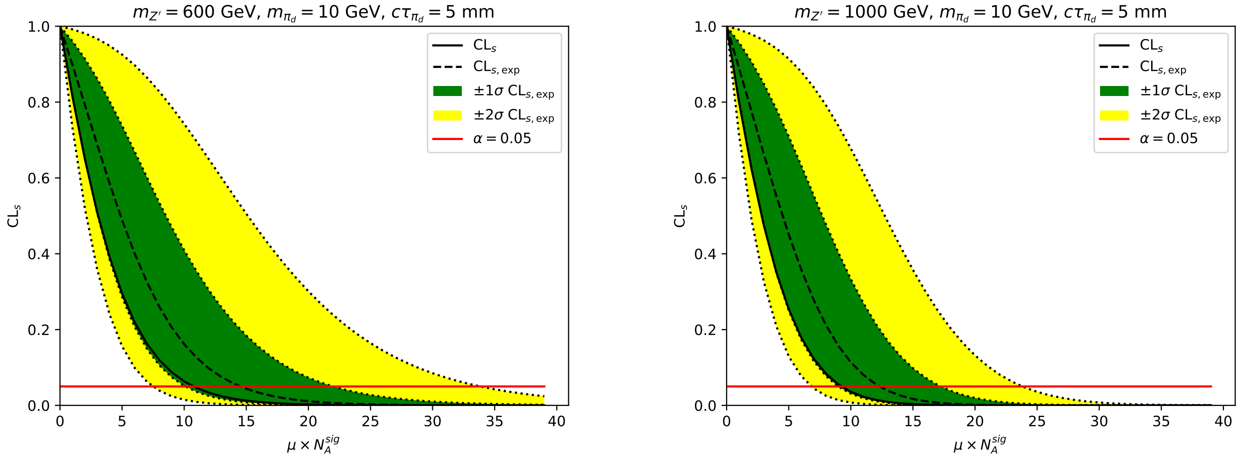
Post-fit	$N_{\mathrm{A}}^{\mathrm{bkg}}$	$N_{\mathrm{B}}^{\mathrm{bkg}}$	$N_{\mathrm{C}}^{\mathrm{bkg}}$	$N_{\mathrm{D}}^{\mathrm{bkg}}$	$\mu \times N_{\mathrm{A}}^{\mathrm{sig}}$
$m_{Z'} = 0.6$ TeV	15 ± 3	41 ± 12	32 ± 9	86 ± 29	0.0 ± 3.7
$m_{Z'} = 0.8$ TeV	15 ± 3	41 ± 12	32 ± 9	86 ± 29	0.0 ± 3.6
$m_{Z'} = 1$ TeV	15 ± 3	41 ± 12	32 ± 9	86 ± 29	0.0 ± 3.4

Table 5.19: Post-fit expected yields for signals corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm.

For the same signals, the observed and expected upper limits (ULs) on $\mu \times N_{\mathrm{A}}^{\mathrm{sig}}$ are presented in Table 5.20. The observed and expected CL$_s$ as a function of the number of signal events in the SR is shown in Figure 5.28. These upper limits on $\mu \times N_{\mathrm{A}}^{\mathrm{sig}}$ are in the same order of magnitude between the different signals and differ mainly on the ± 1 and $\pm 2\sigma$ bands, reflecting the relative systematic uncertainties.

ULs on $\mu \times N_{\mathrm{A}}^{\mathrm{sig}}$	**Observed**	**Expected**	-2σ	-1σ	$+1\sigma$	$+2\sigma$
$m_{Z'} = 0.6$ TeV	10.8	14.6	7.5	10.1	22.0	33.8
$m_{Z'} = 0.8$ TeV	9.6	12.7	7.0	9.3	17.9	25.0
$m_{Z'} = 1$ TeV	9.3	12.3	6.8	9.0	17.3	23.8

Table 5.20: Observed and expected ULs on the number of signal events in the SR for signals corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm.



(a)

(b)

Figure 5.28: Observed and expected CL$_s$ as a function of the number of signal events in the SR for two different signals. The red line indicates the threshold at 0.05.

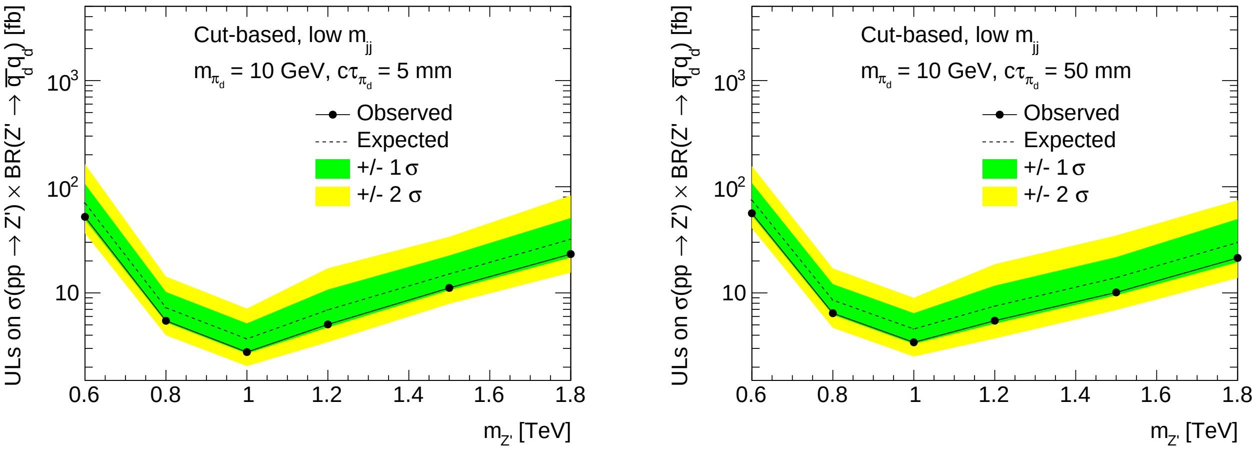
The corresponding ULs on $\sigma(pp \to Z') \times \mathrm{BR}(Z' \to \bar{q}_d q_d)$ are presented in Table 5.21. They have been computed by dividing the upper limits on $\mu \times N_{\mathrm{A}}^{\mathrm{sig}}$ by $\mathcal{L}_{\mathrm{int}} \times f_{\mathrm{eff}} \times k \times A$, with $\mathcal{L}_{\mathrm{int}}$ the integrated luminosity equal to 51.8 fb^{-1}, f_{eff} the efficiency filter, k the $k-$factor (cf section 5.2.2) and A the signal acceptance as evaluated with the MC generated events.

It is possible to see that the UL set on the signal corresponding to $m_{Z'} = 600$ GeV is significantly higher than the ones set on higher masses, while the ULs on the number of signal events in the SR shown in Table 5.20 are of the same order of magnitude. It is a direct consequence of the selection $p_T > 250$ GeV that is applied to the trigger-matched jet, reducing particularly the acceptance of $m_{Z'} = 0.6$ TeV with respect to higher masses. Moreover, because the number of observed events in the SR is lower than the expected one, the observed ULs are lower than the expected ones.

ULs on $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$ [fb]	**Observed**	**Expected**	-2σ	-1σ	$+1\sigma$	$+2\sigma$
$m_{Z'} = 0.6$ TeV	52.1	70.3	36.4	48.8	106.2	163.5
$m_{Z'} = 0.8$ TeV	5.5	7.3	4.0	5.3	10.2	14.3
$m_{Z'} = 1$ TeV	2.8	3.7	2.0	2.7	5.2	7.1

Table 5.21: Observed and expected ULs on $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$ for signals corresponding to $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm.

The 95% CL ULs on $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$ as a function of $m_{Z'}$ for $c\tau_{\pi_d} = 5,\ 50$ mm, and as a function of $c\tau_{\pi_d}$ for $m_{Z'} = 600$ GeV are finally presented in Figure 5.29 and 5.30 respectively.



(a) (b)

Figure 5.29: Observed and expected upper limits on $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$ as a function of $m_{Z'}$ for (a) $c\tau_{\pi_d} = 5$ mm and (b) $c\tau_{\pi_d} = 50$ mm.

The effect of the m_{jj} selection is directly observable in Figures 5.29.a and 5.29.b through the increase of the upper limits from $m_{Z'} = 1.2$ to 1.8 TeV. As it will be explained in the next section, the low-m_{jj} results will be combined with the high-m_{jj} ones in the final analysis results, allowing to set consistent exclusion limits across all the considered range in $m_{Z'}$.

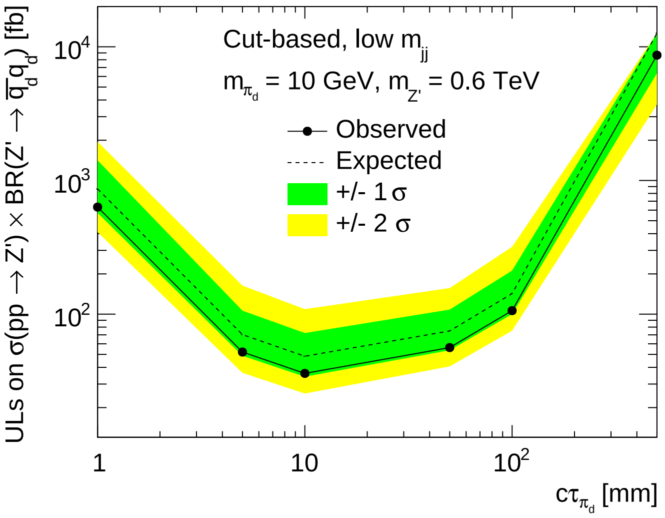


Figure 5.30: Observed and expected upper limits on $\sigma(pp \to Z') \times \mathrm{BR}(Z' \to \bar{q}_d q_d)$ as a function of $c\tau_{\pi_d}$ for $m_{Z'} = 600$ GeV.

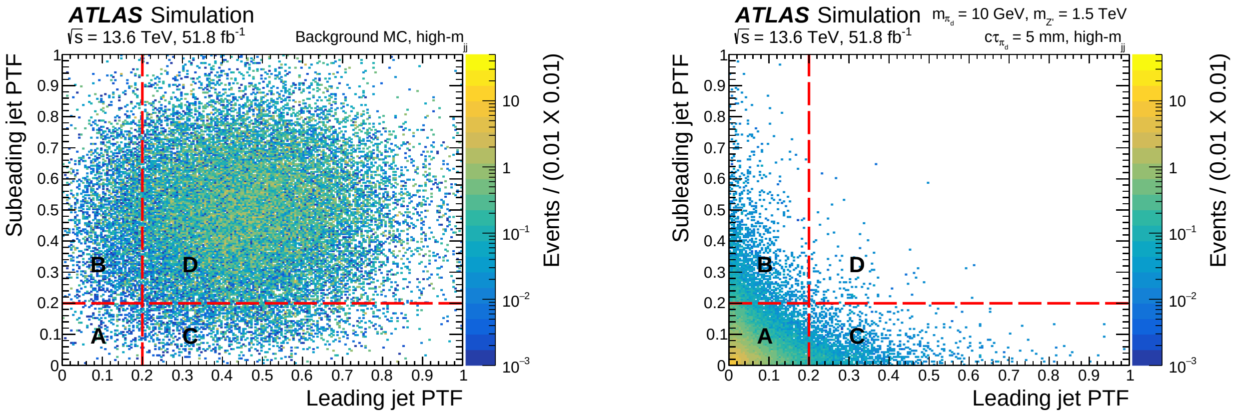
Regarding the upper limits obtained as a function of $c\tau_{\pi_d}$ at fixed $m_{Z'}$ and m_{π_d}, it is possible to notice that the sensitivity is reduced at both low and high $c\tau_{\pi_d}$ values, i.e $c\tau_{\pi_d} \sim 1$ mm and $c\tau_{\pi_d} \sim 500$ mm respectively. In the first case, it comes from the fact that signal events corresponding to lower $c\tau_{\pi_d}$ values are expected to contain jets with larger PTF than the ones corresponding to higher $c\tau_{\pi_d}$, and thus are more subject to be rejected by the trigger and the selection on the PTF. In the second case, this loss of sensitivity is due to the reduced track reconstruction efficiency at higher radius in the inner detector : the secondary vertex reconstruction efficiency is thus also reduced, such as the efficiency of the selection on N_{vtx}.

5.10 Analysis signal regions

The selections that defined the high-m_{jj} cut-based signal region are shown in Figure 5.31 [57]. The background contribution is estimated using the ABCD method with the leading and sub-leading jet PTF defining the four regions, as shown in Figure 5.32.a and 5.32.b for MC background and signal events respectively [57].

Preselection	Low-m_{jj}	High-m_{jj}
leading jet p_T	–	> 520 GeV
subleading jet p_T	–	> 300 GeV
trigger-matched jet	$p_T > 250$ GeV, PTF < 0.04	–
leading and subleading jet PTF	< 1.0	
leading and subleading jet ΔPTF	< 0.4	
leading and subleading jet N_{vtx}	≥ 1	
leading and subleading jet ECF2/p_T	–	> 40 GeV
m_{jj}	< 1 TeV	> 1 TeV
number of jets	≥ 2	
Signal Region		
leading and subleading jet N_{subjet}	≥ 3	–
leading and subleading jet PTF	–	< 0.2

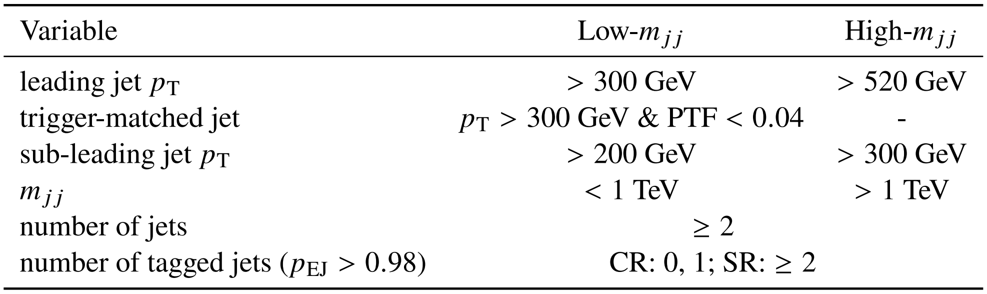
Figure 5.31: Selections defining the low and high-m_{jj} cut-based signal regions [57].



(a) (b)

Figure 5.32: ABCD plane in the high-m_{jj} channel for (a) MC background and (b) signal events corresponding to $m_{Z'} = 1.5$ TeV, $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ mm [57].

On the other hand, the selections defining the ML-based signal regions are shown in Figure 5.33 [57]. After the different pre-selections applied, the SRs are defined by requiring at least two jets tagged as emerging jets, i.e having a classification score (cf Figure 5.22) higher than 0.98.



Variable	Low-m_{jj}	High-m_{jj}
leading jet p_{T}	> 300 GeV	> 520 GeV
trigger-matched jet	p_{T} > 300 GeV & PTF < 0.04	-
sub-leading jet p_{T}	> 200 GeV	> 300 GeV
m_{jj}	< 1 TeV	> 1 TeV
number of jets	≥ 2	
number of tagged jets (p_{EJ} > 0.98)	CR: 0, 1; SR: ≥ 2	

Figure 5.33: Selections defining the low and high-m_{jj} ML-based signal regions [57].

To estimate the background, the ML-based approach rely on a data-driven method based on mis-tag rate. The probability that a given background jet will be mis-tagged as an emerging jet is determined as a function of the jet p_T and PTF. It is directly computed in data in control regions defined with events having strictly less than two jets tagged.

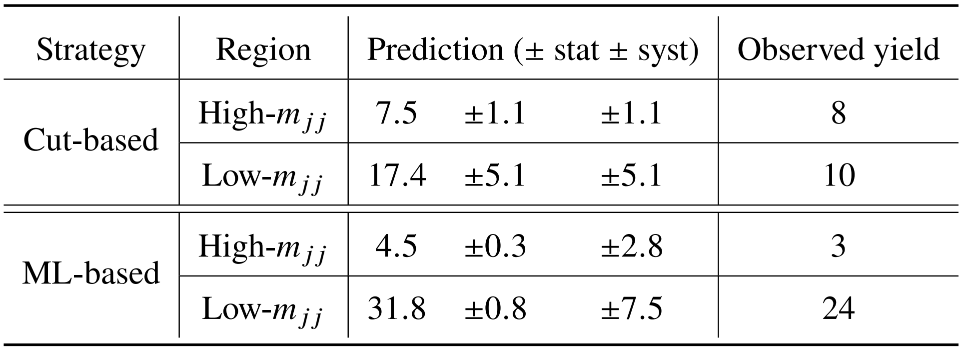
Then, the probabilities to mis-tag exactly zero, one or at least two background jets for a given event are computed using the formulas :

$$\begin{aligned} P(0 \text{ tag}|\text{event}) &= \prod_{i=1}^{n_{\text{jets}}} (1 - P(\text{tag}|j_i)) \\ P(1 \text{ tag}|\text{event}) &= \sum_{i=1}^{n_{\text{jets}}} P(\text{tag}|j_i) \times \prod_{k \neq i} (1 - P(\text{tag}|j_k)) \\ P(\geq 2 \text{ tags}|\text{event}) &= 1 - P(0 \text{ tag}|\text{event}) - P(1 \text{ tag}|\text{event}) \end{aligned}$$

with $P(\text{tag}|j_i)$ the probability to mis-tag the jet i and n_{jets} the number of jets in the event. The background prediction in the SR is then obtained by computing $\sum_{\text{event}} P(\geq 2 \text{ tags}|\text{event})$ with the sum occurring on all the pre-selected events. The way the uncertainties on the background estimation are computed is detailed in [57].

5.11 Analysis results

The observed and expected yields in the different SRs of the analysis are shown in Figure 5.34 [57]. It is possible to see that the observed number of events are all in agreement with the background expectations within the total uncertainties.



Strategy	Region	Prediction (± stat ± syst)			Observed yield
Cut-based	High-m_{jj}	7.5	±1.1	±1.1	8
	Low-m_{jj}	17.4	±5.1	±5.1	10
ML-based	High-m_{jj}	4.5	±0.3	±2.8	3
	Low-m_{jj}	31.8	±0.8	±7.5	24

Figure 5.34: Observed and expected yields in the different SRs including the statistical and systematic uncertainties on the prediction. The predictions in the cut-based regions correspond to the ones obtained with $N_{\text{B}} \times N_{\text{C}}/N_{\text{D}}$ [57].

In order to compare the sensitivity of the two approaches, two different statistical interpretations are performed, one for the cut-based and one for the ML-based strategy. Moreover, for each interpretation, the low and high-m_{jj} regions are combined in the likelihood function.

Concerning the cut-based analysis, the global likelihood is built as :

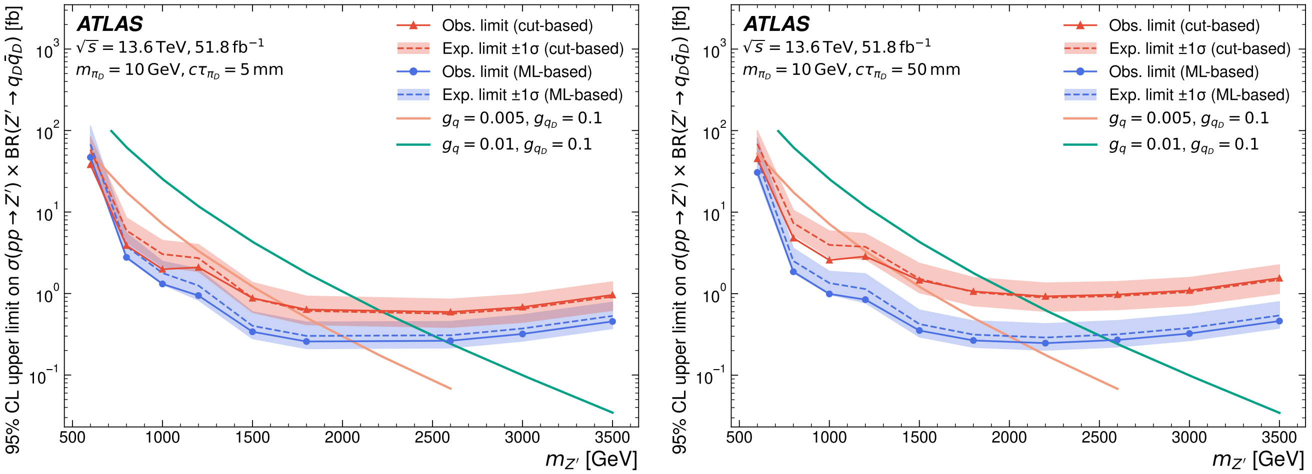
$$\mathcal{L}(\mu, \theta) = \prod_{r \in \{\text{low-}m_{jj},\ \text{high-}m_{jj}\}} \prod_{\text{X} \in \{\text{A, B, C, D}\}} \mathcal{P}(N_{\text{X}(r)} | N^{\text{exp}}_{\text{X}(r)}(\mu, \theta)) \times \prod_i \mathcal{G}(\theta_i) \tag{5.11}$$

with $N_{\text{X}(r)}$ and $N^{\text{exp}}_{\text{X}(r)}$ the observed and expected number of events in the region X in the channel r (either low or high-m_{jj}), and with the gaussian terms incorporating the uncertainties both from low and high-m_{jj} channels. The ABCD method estimation is performed separately in the two channels. Therefore, there are three free parameters $\mu_B^{(r)}$, $\tau_B^{(r)}$ and $\tau_C^{(r)}$ for each channel r. Finally, a common signal strength μ between the two channels is considered in the likelihood, leading to seven free parameters in total. Upper limits at 95 % confidence level on the signal strength and thus on $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$ have been computed using the CL$_s$ method as described in section 5.9.2.

In Figure 5.35 [57], the observed and expected ULs from the cut-based (ML-based) approach are shown in red (blue) as a function of $m_{Z'}$ for $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 5$ and 50 mm. The ML-based approach sets the strongest exclusions except for the signal corresponding to $m_{Z'} = 0.6$ TeV and $c\tau_{\pi_d} = 5$ mm. The reason is that the tagging algorithm used in the ML-based strategy is considering an important number of features, including their possible correlations, for the tracks associated to the jet (cf Figure 5.36 [57]) in order to determine the classification score. Since the discriminating power of the tagging algorithm is degraded at low p_T, the low-m_{jj} ML-based analysis employs a p_T threshold at 300 GeV instead of 250 GeV, which explains why the cut-based approach is able to perform better than the ML-based one for the $m_{Z'} = 0.6$ TeV signal. For other signals, even if PTF and N_{vtx} are highly discriminant variables between background and signal events, the sensitivity achieved in the cut-based approach with selections on these variables is not as great as the one obtained in the ML-based approach.

Moreover, it is possible to notice visually which channel between the low and high-m_{jj} ones is predominant during the establishment of the cross-section upper limits. Indeed in the low-m_{jj} channel, the observed number of events is around 1σ below the expected yield in both cut-based and ML-based approaches, while these numbers are closer in the high-m_{jj} channels. In this way, if the low-m_{jj} channel is predominant, the observed limit would be 1σ below the expected one while if it is the high-m_{jj} channel that is preponderant, the observed and expected limits would be closer.

At a fixed value of $c\tau_{\pi_d}$, the upper limits are slightly increasing with $m_{Z'}$. It comes from the fact that higher values of $m_{Z'}$ lead to higher boost for the produced π_d. The π_d are thus subject to decay at higher radius in the inner detector where the track and secondary vertex reconstruction efficiencies are degraded compared to lower radius.



(a) $c\tau_{\pi_d} = 5$ mm

(b) $c\tau_{\pi_d} = 50$ mm

Figure 5.35: Upper limits at 95% confidence level on $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$ as a function of $m_{Z'}$ for $m_{\pi_d} = 10$ GeV and (a) $c\tau_{\pi_d} = 5$ mm and (b) $c\tau_{\pi_d} = 50$ mm. The theoretical cross-section values are shown as solid lines for different g_q and g_{q_d} [57].

In Figure 5.37 [57], the observed and expected ULs are shown as a function of $c\tau_{\pi_d}$ for $m_{\pi_d} = 10$ GeV and $m_{Z'} = 0.6$, 1.5 and 3 TeV. For the cut-based approach, the behavior of the upper limits as a function of $c\tau_{\pi_d}$ is similar to the one observed in Figure 5.30 for the same reasons. Concerning the ML-based approach for $m_{Z'} = 1.5$ and 3 TeV, the limits are similarly degraded at high $c\tau_{\pi_d}$ values due to the reduced track reconstruction efficiency, reducing the discriminating power of the tagging algorithm. Moreover, contrary to the cut-based approach, the limits are roughly constant across $c\tau_{\pi_d}$ values between 1 and 100 mm : indeed, the high-m_{jj} channel from the ML-based approach

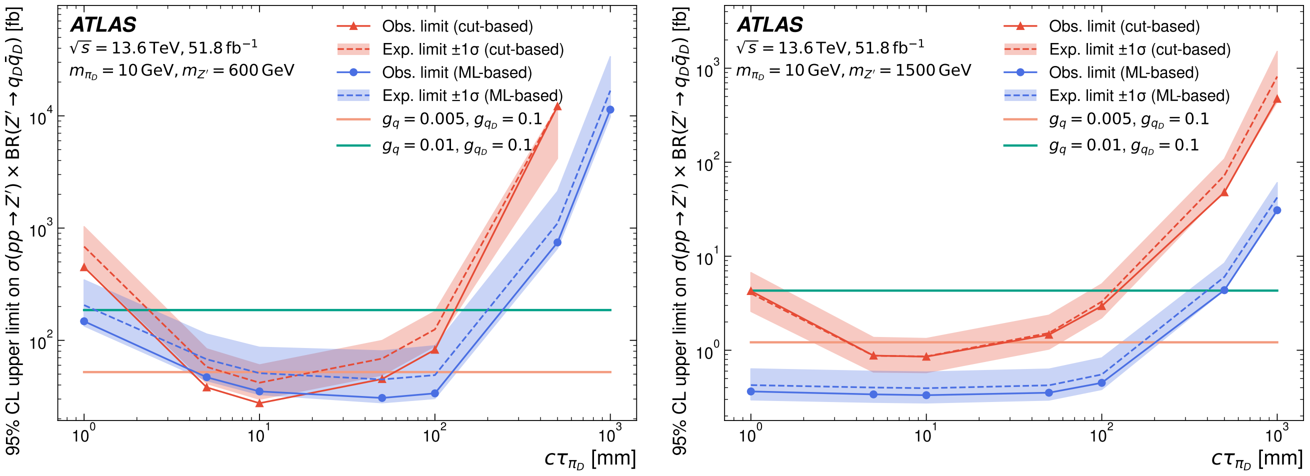
Input	Description
Jet η	Jet pseudorapidity
d_0	Track closest distance to PV in transverse plane
$z_0 \sin(\theta)$	Track closest distance to PV in longitudinal plane
$\Delta\phi$	Azimuthal angle of the track, relative to the jet ϕ
$\Delta\eta$	Track pseudorapidity, relative to jet η
q/p	Track charge over momentum
$\sigma(\phi)$	Uncertainty in track ϕ
$\sigma(\theta)$	Uncertainty in track θ
$\sigma(q/p)$	Uncertainty in track q/p
$d_0/\sigma(d_0)$	signed d_0 significance
$z_0/\sigma(z_0)$	signed z_0 significance
$N_{\text{PIX hits}}$	Number of Pixel hits per track
$N_{\text{SCT hits}}$	Number of SCT hits per track
$N_{\text{IBL hits}}$	Number of innermost pixel layer hits
$N_{\text{PIX shared}}$	Number of Pixel shared hits
$N_{\text{SCT shared}}$	Number of SCT shared hits

Figure 5.36: List of track features used as inputs in the emerging jet tagging algorithm. The entry «Jet η» is common to all the tracks that are associated to the same jet. A «shared hit» corresponds to a hit that is common between different tracks [57].

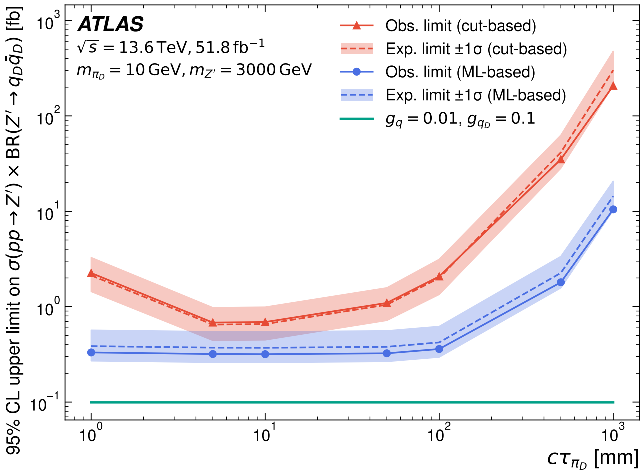
can be sensitive to signal corresponding to very low $c\tau_{\pi_d}$ since there is no selection on PTF in this approach [57].

In Figure 5.38 [57], the ULs are shown as a function of m_{π_d} for $m_{Z'} = 1.5$ TeV and $c\tau_{\pi_d} = 5$, 50 and 500 mm. There is no strong dependence to the dark pion mass except for the signal corresponding to $c\tau_{\pi_d} = 500$ mm. Indeed, at a fixed $m_{Z'}$ and $c\tau_{\pi_d}$, the lighter dark pions are more boosted and thus decay at higher radius in the inner detector with respect to the heavier ones. For the $c\tau_{\pi_d} = 500$ mm signal, the dark pion decays can even occur outside the inner detector, and thus lead to a loss of sensitivity.

By assuming fixed value for g_q and g_{q_d}, the theoretical value of $\sigma(pp \to Z') \times \text{BR}(Z' \to \bar{q}_d q_d)$ can be computed as a function of the model parameters. Then, a signal is excluded if the corresponding theoretical cross-section is greater than the upper limit. In this way, it is possible to derive exclusion contours in the $(m_{Z'},\ c\tau_{\pi_d})$ 2D plane as shown in Figure 5.39 [57], assuming $g_q = 0.01$, $g_{q_d} = 0.1$ and $m_{\pi_d} = 10$ GeV. The signals corresponding to $m_{Z'}$ and $c\tau_{\pi_d}$ values that are within the red (blue) contour are thus excluded by the cut-based (ML-based) approach. Under these assumptions on g_q, g_{q_d} and m_{π_d}, the cut-based (ML-based) approach allows to exclude Z' masses up to 2200 (2550) GeV for $c\tau_{\pi_d} = 5$ mm for instance. Moreover for $m_{Z'} = 1.0$ TeV, the cut-based (ML-based) approach excludes π_d proper decay length between 1.5-200 (1-500) mm.



(a) $m_{Z'} = 0.6$ TeV (b) $m_{Z'} = 1.5$ TeV



(c) $m_{Z'} = 3.0$ TeV

Figure 5.37: Upper limits at 95% confidence level on $\sigma(pp \to Z') \times \mathrm{BR}(Z' \to \bar{q}_d q_d)$ as a function of $c\tau_{\pi_d}$ for $m_{\pi_d} = 10$ GeV and (a) $m_{Z'} = 0.6$ TeV, (b) $m_{Z'} = 1.5$ TeV and (c) $m_{Z'} = 3.0$ TeV [57].

Finally, exclusions contours in the 2D plane defined by $(g_q,\ g_{q_d})$ can be obtained for a given signal characterized by a set of $m_{Z'}$, m_{π_d} and $c\tau_{\pi_d}$ values. Indeed, the theoretical cross-section can be computed for various values of g_q and g_{q_d}, and the ones that lead to cross-section higher than the upper limit are thus excluded, as shown in Figure 5.40.a (5.40.b) [57] for the ML-based approach for the signal corresponding to $m_{Z'} = 1.5$ (2.6) TeV, $m_{\pi_d} = 10$ GeV and $c\tau_{\pi_d} = 50$ mm.

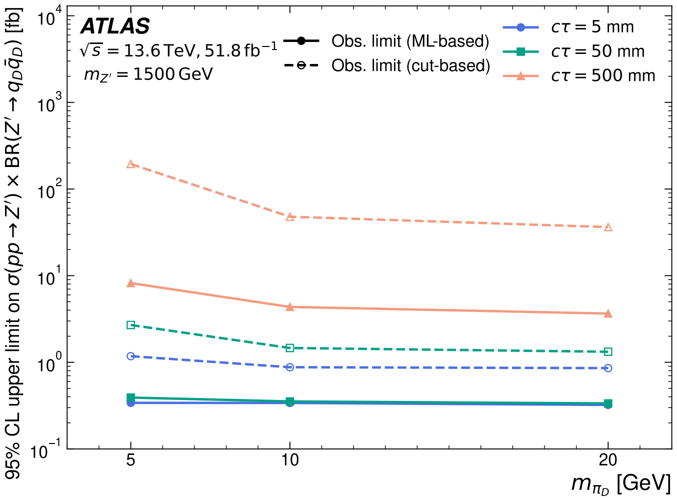


Figure 5.38: Upper limits at 95% confidence level on $\sigma(pp \to Z') \times \mathrm{BR}(Z' \to \bar{q}_d q_d)$ as a function of m_{π_d} for $m_{Z'} = 1.5$ TeV and $c\tau_{\pi_d} = 5$, 50 and 500 mm [57].

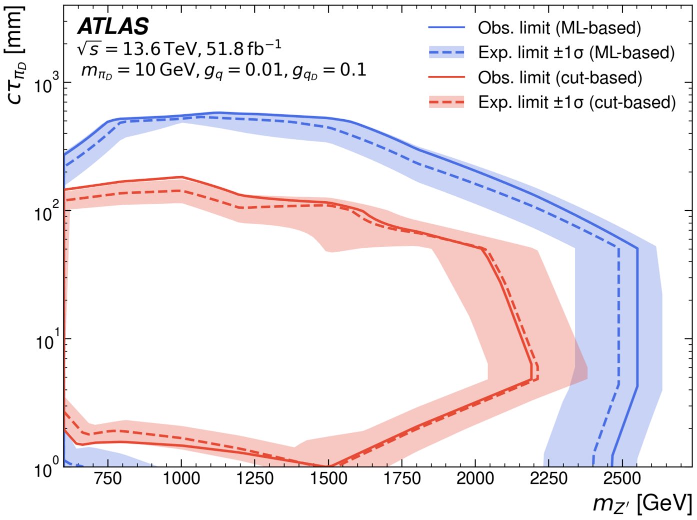
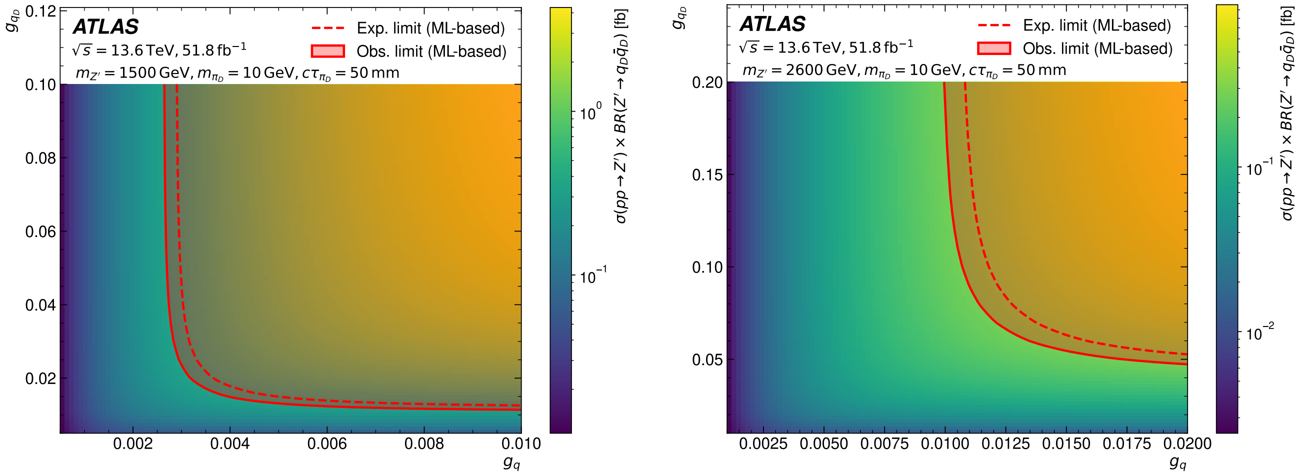


Figure 5.39: Excluded values of $m_{Z'}$ and $c\tau_{\pi_d}$ at 95% confidence level assuming $g_q = 0.01$ and $g_{q_d} = 0.1$ [57].

For example for $m_{Z'} = 1.5$ (2.6) TeV, assuming $g_{q_d} > 0.03$ (0.10), the ML-based approach excludes values of g_q higher than 0.003 (0.01). These limits can be compared to the ones set by previous resonance searches performed in ATLAS that are shown in Figure 5.41 [85] and 5.42 [86]. For instance, the dijet resonance analysis that was searching for new resonances in m_{jj} distributions using the Run-2 data set an upper limit on g_q about 0.06 (0.13) for $m_{Z'} = 1.5$ (2.6) TeV [85], i.e about 20 (13) times higher than the limit sets by the ML-based approach. It thus demonstrates the important gain in sensitivity to the Z' mediator achieved in this analysis with respect to previous ones.



(a) $m_{Z'} = 1.5$ TeV

(b) $m_{Z'} = 2.6$ TeV

Figure 5.40: Excluded values of g_q and g_{q_d} at 95% confidence level for $m_{\pi_d} = 10$ GeV, $c\tau_{\pi_d} = 50$ mm, (a) $m_{Z'} = 1.5$ TeV and (b) $m_{Z'} = 2.6$ TeV. The observed exclusions correspond to the shaded red areas [57].

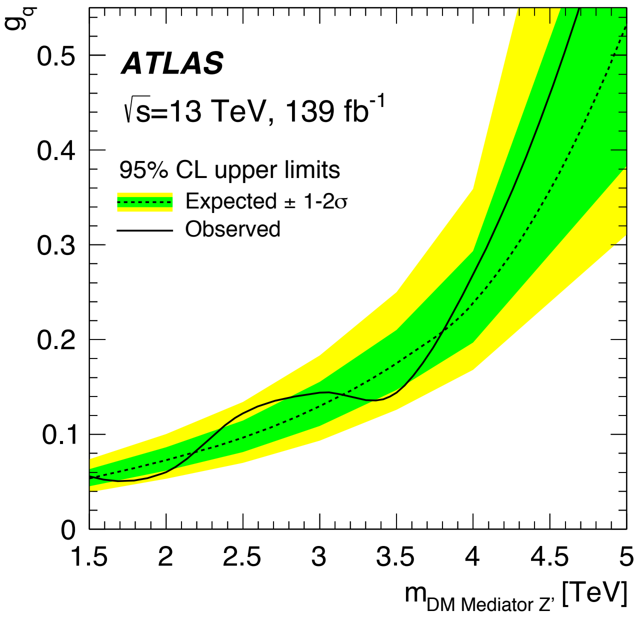


Figure 5.41: Upper limits on the Z' coupling constant to SM quarks g_q at 95% confidence level as a function of $m_{Z'}$ obtained by the ATLAS dijet resonance search performed with Run-2 data [85].

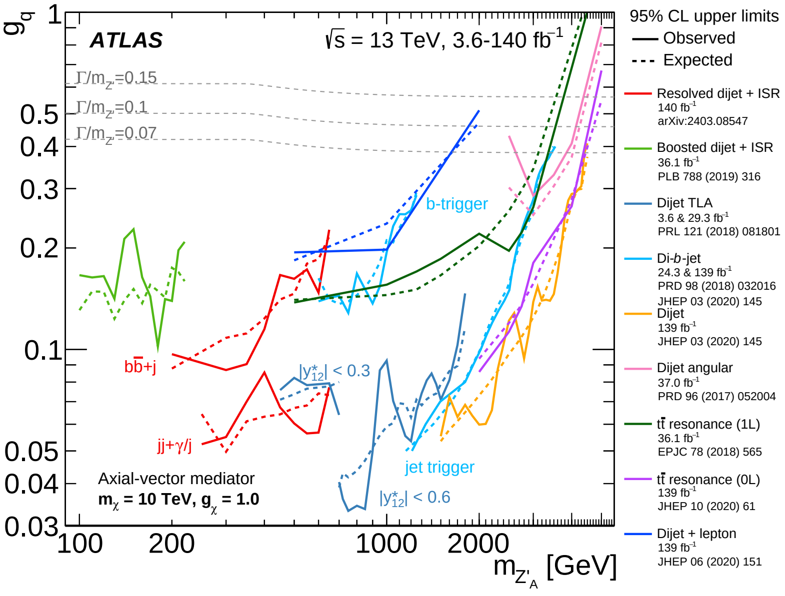


Figure 5.42: Upper limits on g_q at 95% confidence level as a function of the mass of a spin-1 mediator coupled to SM quarks $m_{Z'_A}$ for several ATLAS dijet searches [86].

5.12 Conclusion

This section presented a search for emerging jets using Run-3 data collected by ATLAS in 2022 and 2023. It is the first time a heavy Z' mediator has been considered in a search to produce emerging jets via an s-channel mechanism. Indeed, previous searches performed by CMS focused on a scalar mediator, with a final state corresponding to two emerging jets in addition to two SM jets [87, 88].

This analysis is made of two different strategies : a cut-based one that selects events based on track or substructure based jet observables, and a machine-learning based one exploiting a tagging algorithm to discriminate background from emerging jets. Both strategies use data-driven methods to estimate the background expectation : the ABCD method has been used for the cut-based strategy while the ML-based one employs a method based on mis-tag rates computation. Moreover, in order to improve at maximum the sensitivity, each strategy is divided into two orthogonal regions using two different triggers combined to a selection on the dijet invariant mass m_{jj}.

This section focused on the low-m_{jj} cut-based strategy to which I contributed in a significant way, before presenting the final results obtained with all the different channels of the analysis. No excess has been observed in data with respect to the different background expectations, and in comparison to previous LHC searches, the limits that have been set on the different parameters of the model demonstrate a significant gain in sensitivity to processes involving a new s-channel mediator.

In the next section, the first studies that I have made in order to allow the future reinterpretation of the low-m_{jj} cut-based analysis will be presented.

Chapter 6

Reinterpretation of LHC Dark QCD results

Contents

6.1 Reinterpretation of LHC analyses

The reinterpretation of an LHC analysis consists in exploiting the results of existing ATLAS and CMS analyses that were searching for BSM processes in order to study and test alternative theoretical models. Indeed, a given LHC analysis can only focus on a reduced number of theoretical models to interpret the observed data. The reinterpretation allows to re-employ the analysis results and to confront to the data other models that could have been considered : it is thus fundamental in the context of BSM searches. Moreover, results from reinterpretation can guide the elaboration of new physic theories or reveal holes in the coverage of existing analyses in order to motivate new ones.

Several software frameworks exist to reinterpret LHC analyses. This section will focus on the one named «MadAnalysis» [89, 90, 91], since it is the one I used for the results that will be presented in this section. This framework contains a public database [92, 93] that regroups several ATLAS or CMS analyses implemented in the C++ language, and that can be reinterpreted in a straightforward and accessible way.

For each implemented analysis, there is thus a corresponding C++ code that reproduce the global analysis strategy, i.e the reconstruction of the exploited kinematic observables and the definition of the different channels of selection. Moreover, the experimental results in terms of observed and expected number of events in each channels are also associated to the implemented code.

To reinterpret an analysis with a given theoretical physical process, signal events are first generated at the truth level (i.e after the hadronization but before any interaction with the detector) using a particle generator. To simulate the interaction of the final state particles with the ATLAS or CMS detector, MadAnalysis relies either on Delphes 3 [94] or on a simplified and fast simulation (SFS) through the use of smearing and efficiency functions for object reconstruction and tagging for instance [95]. The analysis code can then be applied to the events after the simulation of the detector response, and the number of signal events in the different SRs can be computed. The latter can then be compared to the observed and expected number of events in order to derive upper limits on the signal cross-section.

During my thesis, I have contributed to the implementation of several LHC analyses related to final state containing jets that can be used to constrain dark QCD theories. These analyses are :

- the ATLAS Run-2 search for new resonances decaying into a pair of jets [85],
- the ATLAS Run-2 search for non-resonant production of semi-visible jets [96],
- and the ATLAS Run-2 search for resonant production of a pair of dark quarks producing a dijet final state (referred as «dark jets» in section 2.4) [97].

Moreover, studies are currently performed to exploit these analyses in addition to other ones that have been already implemented, such as the CMS Run-2 search for mono-jet final state [98], in order to constrain the dark QCD parameter phase space.

6.2 Implementation of LHC analyses within the MadAnalysis database

There are several steps before implementing an analysis. The first one is to write the analysis code based on information that can be found in the published article. The code must reproduce the same event selection as in the analysis, i.e with the same pre-selections and selections on high-level observables that are applied to objects such as jets, leptons or photons and that defined the different SRs.

Concerning the analysis I have implemented, I have exploited the simplified simulation of the detector response. The second step is thus to configure this fast simulation. Depending on the analysis final state, it can be done by setting the jet algorithm that is used in the analysis or to configure the reconstruction efficiencies for the different physic objects (jets, leptons, etc), which can be parametrized as a function of the p_T and η of the object. The object will be thus either kept or thrown away according to the corresponding efficiency. The efficiencies of identify b-jets for instance have also to be set if tagging algorithms are employed in the analysis. Moreover, smearing can be applied to the truth jet or lepton energy : it modifies the energy of the object according to a gaussian distribution that is centered around the truth value and having a width that can be configured.

Then, the next step is to validate the code that has been written, i.e to be able to reproduce in a satisfying way the analysis results concerning the benchmark models that have been tested in the analysis. The events corresponding to these models have thus to be generated with the same parameters as the ones used by ATLAS or CMS. The MadAnalysis code can then be executed on these events, and the cutflows or the histograms obtained can be compared to the ones from the analysis results if they are provided. Finally, by retrieving the number of observed and expected events in the different SRs, the upper limits on the signal cross-section can be computed with MadAnalysis and compared to the ones from the published results.

Having a good agreement between the MadAnalysis results and the ones from ATLAS or CMS allows to validate the implementation of the analysis, that can then be added to the MadAnalysis database.

6.2.1 ATLAS search for new resonances in m_{jj} distributions

Description of the analysis

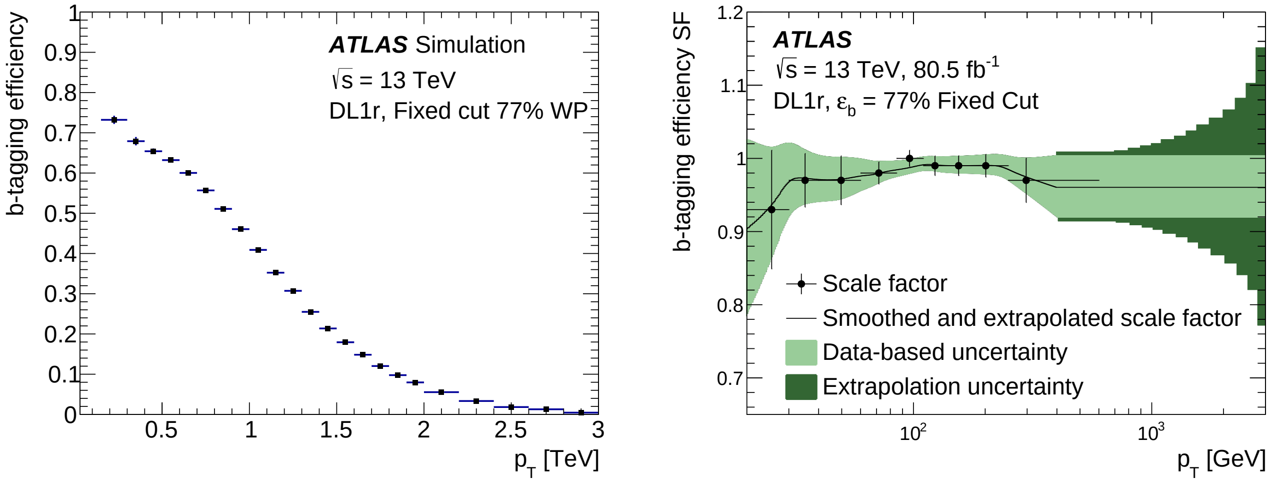
This analysis considers a dijet final state and searches for new resonances from new BSM heavy particles in the dijet system invariant mass m_{jj}. It considers Run-2 data collected between 2015 and 2018 that corresponds to an integrated luminosity of 139 fb^{-1}. The strategy of the analysis consists in examining the m_{jj} distributions in the different SRs and searching for local excesses above a background contribution estimated with data driven techniques [85].

Given that no significant excess has been observed in data, the results of the analysis have been translated into upper limits at 95% confidence level on the signal cross-section for several theoretical models.

The events are first selected with a trigger requiring at least one jet satisfying $p_T > 420$ GeV. The jets are reconstructed with topological clusters as constituents that are clustered with the anti-k_t algorithm with a radius parameter of $R = 0.4$.

The events are required to have at least two jets with $p_T > 150$ GeV with an azimuthal separation between the two leading jets satisfying $|\Delta\phi(jj)| > 1.0$. The events are then classified into three regions : an inclusive region, a one-b-tagged region ($1b$) corresponding to final state with at least one jet tagged as a b-jet, and a two-b-tagged region ($2b$) with exactly two jets tagged as b-jets.

The jets containing b-hadrons are identified as b-jet using the deep-learning neutral network DL1r at the 77% efficiency working point, which considers features from track impact parameters and displaced vertices. The tagging efficiencies of the DL1r network as a function of the jet p_T are presented in Figure 6.1.a [85]. Moreover, in order to correct for differences between MC and data events, a scale factor that depends on the jet p_T is applied to simulated events. The way the scale factors have been extracted is reported in [99], and Figure 6.1.b presents the values of this scale factor as a function of the jet p_T [85].



(a)

(b)

Figure 6.1: (a) b-tagging efficiency of the DL1r algorithm as a function of the jet p_T and (b) corresponding simulation-to-data scale factor as a function of the jet p_T [85].

In the b-tagged categories, the two leading jets must satisfy $|\eta| < 2.0$. A selection on the centrality $y^* = 0.5 \times (y_1 - y_2)$, with y_i the rapidity of the jet i, is applied, in addition to a lower bound on the invariant mass m_{jj}. In the inclusive region, the selections on these two variables are either $|y^*| < 0.6$ & $m_{jj} > 1100$ GeV or $|y^*| < 1.2$ & $m_{jj} > 1717$ GeV, depending on the targeted signal. In the b-tagged regions, the selection is $|y^*| < 0.8$ & $m_{jj} > 1133$ GeV.

The interpretation of the results has been performed considering various signal models decaying into a pair of quarks or gluons such as excited quarks q^*, leptophobic dark matter Z' mediator, heavy charged W' gauge boson, etc.

Implementation and validation

To implement this analysis, I have written an analysis code that reproduces the selections of the three different SRs. I have also configured the b-tagging efficiencies as a function of the jet p_T considering the Figures 6.1.a and 6.1.b by multiplying the tagging efficiencies to the simulation-to-data scale factors.

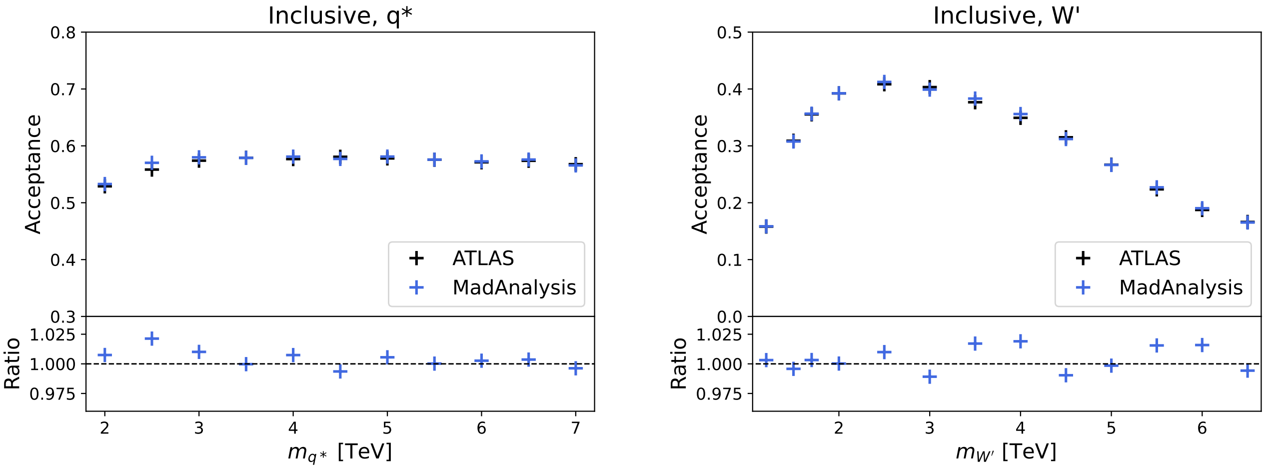
To validate the implementation, I have considered several signal processes :

- for the inclusive region, the productions of an excited quark [100] and a W' boson [101] through $q\bar{q} \to q^*$ with $q^* \to qg$, and $q\bar{q} \to W'$ with $W' \to q\bar{q}$,
- for the 1b region, the production of an excited b-quark through $q\bar{q} \to b^*$ with $b^* \to bg$, $b\gamma$, bZ, tW with a branching ratio to bg equal to 85% [100],

- for the $2b$ region, the production of a sequential standard model Z' boson through $q\bar{q} \to Z'$ with $Z' \to b\bar{b}$ [102].

For each signal, I have generated 20k events with Pythia 8.308 [78] for different masses of the BSM particles, using the same parameters as the ones employed by ATLAS.

I have computed the acceptance for each signal as a function of the particle mass and compared the obtained values to the ones published by ATLAS, as shown in Figures 6.2 and 6.3. The agreement with ATLAS is excellent, especially for the signals in the inclusive and $1b$ regions where the relative differences between the MadAnalysis and ATLAS values are below 3%. For the signal in the $2b$ region, the relative differences are at maximum about 15%, which remains an acceptable results.

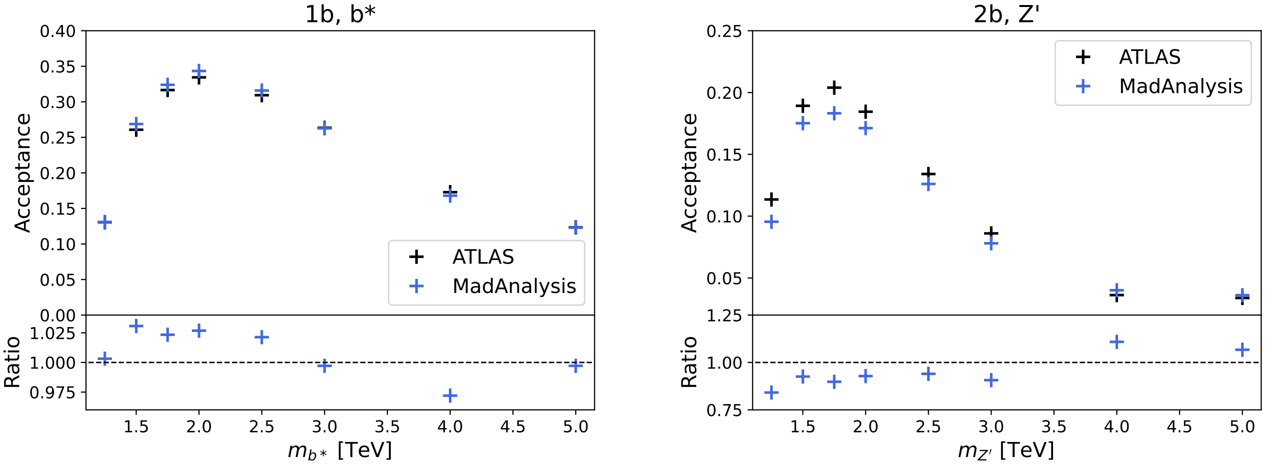


(a) (b)

Figure 6.2: Acceptance in the inclusive SR for the signal corresponding to (a) an excited quark and (b) a W' boson.

In order to compute the upper limits on the signal cross-section, the numbers of observed and expected events in each m_{jj} bin, in addition to the uncertainties on the background contributions, for the three different SRs have been obtained from ATLAS published results.

For each signal, the upper limit on the cross-section can be computed separately in each m_{jj} bin with the procedure described in section 5.3 in [91]. However, selecting the upper limit obtained in the m_{jj} bin that demonstrates the best sensitivity is not optimal to reproduce the ATLAS limits. Instead, a combination of all the m_{jj} bins in a given SR (inclusive, $1b$ or $2b$) has been exploited through the use of a simplified likelihood [103, 104]. It is done by configuring the covariance matrix that encodes the correlations between the expected background contributions in the different m_{jj} bins of a given SR. Since ATLAS does not provide such information concerning this analysis, it has been decided to set a covariance matrix as being diagonal, i.e corresponding to uncorrelated background contributions among the different bins.

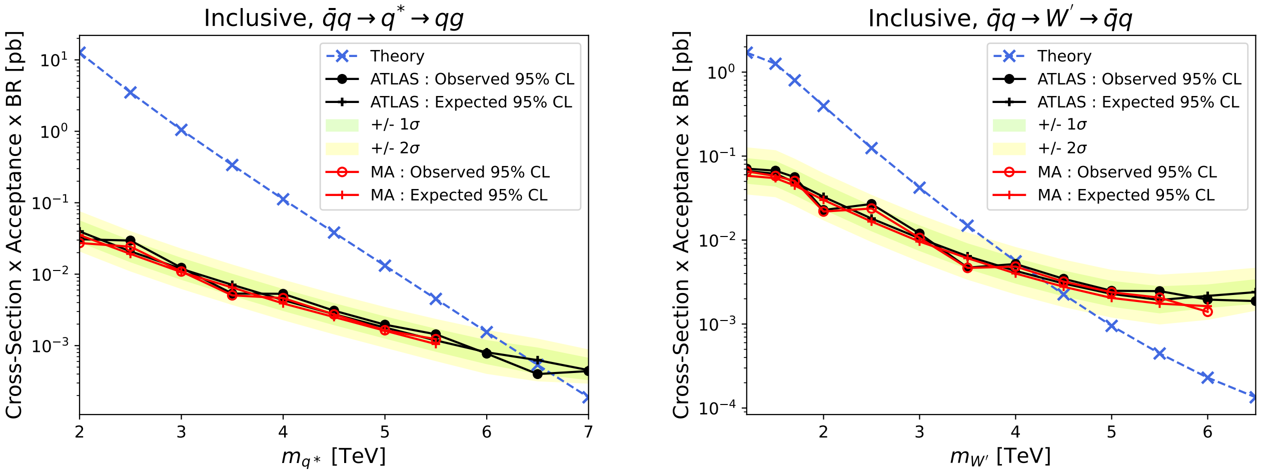


(a) (b)

Figure 6.3: Acceptance in the (a) 1b SR for the signal corresponding to an excited b-quark and in the (b) 2b SR for the signal corresponding to a Z' boson.

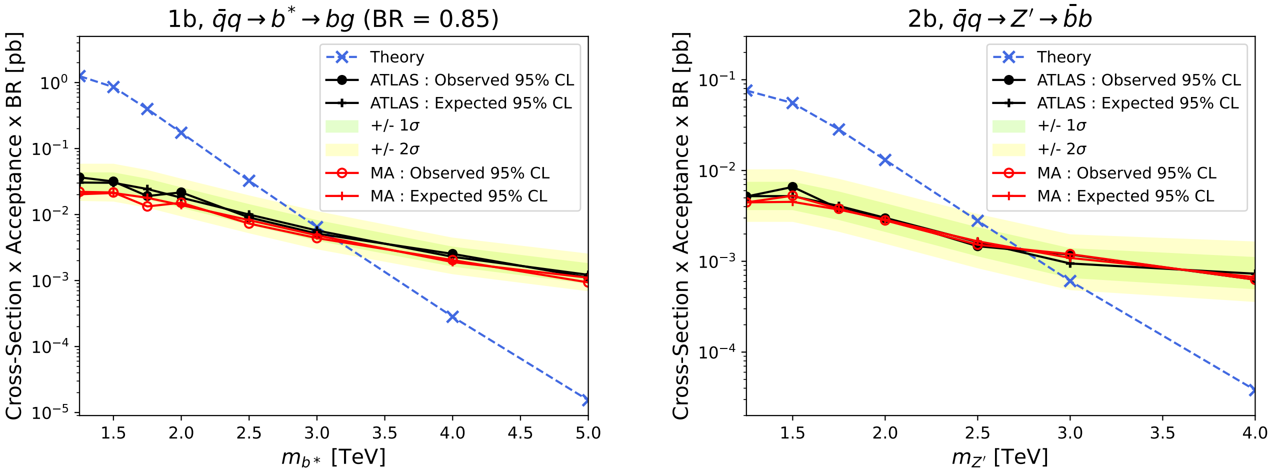
The upper limits obtained for the different signals as a function of the mass of the BSM particles are shown in Figure 6.4 and 6.5. The agreement with the ATLAS limits is globally excellent across the different signals that have been considered.

The analysis has thus been implemented in the MadAnalysis database, and is available for reinterpretation with theoretical models that produce a signature corresponding to a pair of QCD jets.



(a) (b)

Figure 6.4: Upper limits on the cross-section multiplied by the acceptance and the branching ratio into two jets for the signals corresponding to (a) an excited quark and (b) a W' boson. «MA» accounts for MadAnalysis.



(a) (b)

Figure 6.5: Upper limits on the cross-section multiplied by the acceptance and the branching ratio into two jets for the signals corresponding to (a) an excited b-quark and (b) a Z' boson.

6.2.2 ATLAS search for semi-visible jets

Description of the analysis

This analysis was searching for semi-visible jets, corresponding to final states with jets where one of them is aligned with the direction of the missing transverse momentum. It considers the production of semi-visible jets via a t-channel mediator Φ that produces a pair of dark quarks as shown in Figure 6.6 [96]. These dark quarks then undergo parton shower and hadronization in the dark sector to form either stable or unstable dark hadrons, before these latter decay back to the visible sector. The model is parametrized by the mass of the mediator m_Φ, the fraction of stable dark hadrons among all the dark hadrons in the events R_{inv}, and the coupling constant λ between a quark, a dark quark and the mediator. The analysis uses 139 fb^{-1} of data at $\sqrt{s} = 13$ TeV collected by ATLAS between 2015 and 2018.

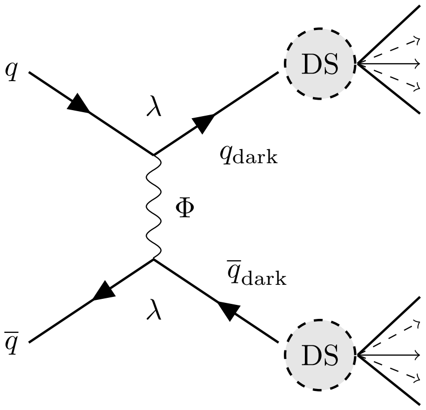


Figure 6.6: Production of semi-visible jets via a t-channel mediator Φ. «DS» denotes the dark shower [96].

The events are required to pass a trigger with a threshold on the missing transverse energy E_T^{miss}. This threshold was equal to 70 GeV in 2015, and reached 110 GeV during the 2016-2018 data-taking period. The jets of the events are built from Particle Flow constituents using the anti-k_t algorithm with a radius of 0.4.

The events must have two reconstructed jets within $|\eta| < 2.8$, with the leading jet satisfying $p_T > 250$ GeV and the other having $p_T > 30$ GeV. Moreover, at least one jet has to be within $\Delta\phi < 2.0$ of the p_T^{miss} direction.

The events with two or more b-jets are rejected, with the identification of b-jets done using the DL1r algorithm at the 77 % efficiency working point similarly as the dijet analysis described previously. Moreover, hadronically decaying τ-lepton candidates are formed thanks to a Recurrent Neural Network algorithm [105] with information from the calorimeters and the inner detector : any event containing such a τ-lepton with $p_T > 20$ GeV and $|\eta| < 2.5$ is rejected. Events with electrons or muons passing $p_T > 7$ GeV are discarded.

The signal region is finally defined with $E_T^{\text{miss}} > 600$ GeV and $H_T = \sum_{\text{jets}} p_T > 600$ GeV. It is divided into 9 bins using two discriminating and uncorrelated variables, which are :

- the p_T balance between the closest (j_1) and farthest (j_2) jets in azimuth from the p_T^{miss} direction, that is defined as $p_T^{\text{bal}} = |\vec{p_T}(j_1) + \vec{p_T}(j_2)|/(|\vec{p_T}(j_1)| + |\vec{p_T}(j_2)|)$,
- and the azimuthal distance between j_1 and j_2 denoted by $|\phi_{\max} - \phi_{\min}|$.

The definition of these bins is shown in Table 6.1. The background contribution is estimated in each of these bins using control regions containing events with muons and the same E_T^{miss} and H_T requirements as the signal region. No excess has been observed in data and exclusion on the signal strength has been done thanks to a simultaneous binned likelihood fit in the signal region and in the control regions, before using the CLs method to set upper limits on the signal cross section at 95% confidence level.

	$\|\phi_{\max} - \phi_{\min}\| < 2.0$	$2.0 < \|\phi_{\max} - \phi_{\min}\| < 2.7$	$2.7 < \|\phi_{\max} - \phi_{\min}\|$
$p_T^{\text{bal}} < 0.6$	Bin 1	Bin 2	Bin 3
$0.6 < p_T^{\text{bal}} < 0.9$	Bin 4	Bin 5	Bin 6
$0.9 < p_T^{\text{bal}}$	Bin 7	Bin 8	Bin 9

Table 6.1: Definition of the 9 bins in the signal region using the two variables p_T^{bal} and $|\phi_{\max} - \phi_{\min}|$.

Implementation and validation

Concerning this analysis, I did not write the analysis code but I contributed to the validation of its implementation. The code reproduces all the selections described previously, and the b-tagging efficiencies that have been configured are the same as the ones employed for the dijet analysis, since it is the same neural network at the same working point that has been employed in the semi-visible jets analysis.

ATLAS provides cutflows and signal yields in the signal region for 6 representative models that correspond to : $m_\Phi = 1$ TeV & $R_{\text{inv}} = 0.6$, $m_\Phi = 1$ TeV & $R_{\text{inv}} = 0.8$, $m_\Phi = 2$ TeV & $R_{\text{inv}} = 0.4$, $m_\Phi = 2$ TeV & $R_{\text{inv}} = 0.6$, $m_\Phi = 3$ TeV & $R_{\text{inv}} = 0.2$ and $m_\Phi = 3$ TeV & $R_{\text{inv}} = 0.4$. For each of these models and using the same parameters as ATLAS, I have generated around 30k events with MadGraph5 aMC v2.6.7 [33] for the hard scattering matrix elements computation, and Pythia 8.244 [30] with the Hidden Valley module for the simulation of the parton shower and hadronization in both visible and dark sectors.

I made comparison between the cutflows obtained with MadAnalysis and the ones published by ATLAS as shown in Tables 6.2 to 6.7. These cutflows show the ATLAS (MadAnalysis) number of events after each cut weighted by $\mathcal{L} \times \sigma$ (with $\mathcal{L}$ the integrated luminosity and σ the signal cross-section computed with $\lambda = 1$) in the column «ATLAS» («MA»). The column «Relative difference» corresponds to $|N_{\text{ATLAS}} - N_{\text{MA}}|/N_{\text{ATLAS}}$, with N the number of weighted events at a given step in the cutflow. The columns «ATLAS ϵ» and «MA ϵ» represent the selection efficiency of a given cut for ATLAS and MadAnalysis respectively.

	ATLAS	**MA**	**Relative difference** [%]	**ATLAS** ϵ [%]	**MA** ϵ [%]
Pre-selections	844520	802004	5		
$\Delta\phi < 2.0$	816341	773773	5	97	96
$p_T^{\text{lead.}} > 250$ GeV	791042	747516	5	97	97
$N_{b\text{-jet}} < 2$	707151	747516	6	89	100
τ jet veto	701538	724219	3	99	97
$E_T^{\text{miss}} > 600$ GeV	101378	108504	7	14	15
$H_T > 600$ GeV	101235	108504	7	100	100

Table 6.2: Cutflow for the model corresponding to $m_\Phi = 1$ TeV & $R_{\text{inv}} = 0.6$.

	ATLAS	**MA**	**Relative difference** [%]	**ATLAS** ϵ [%]	**MA** ϵ [%]
Pre-selections	958068	937324	2		
$\Delta\phi < 2.0$	848395	887855	5	89	95
$p_T^{\text{lead.}} > 250$ GeV	812949	868463	7	96	98
$N_{b\text{-jet}} < 2$	760522	868463	14	94	100
τ jet veto	754784	769612	2	99	89
$E_T^{\text{miss}} > 600$ GeV	159627	164451	3	21	21
$H_T > 600$ GeV	159582	164236	3	100	100

Table 6.3: Cutflow for the model corresponding to $m_\Phi = 1$ TeV & $R_{\text{inv}} = 0.8$.

	ATLAS	**MA**	**Relative difference** [%]	**ATLAS** ϵ [%]	**MA** ϵ [%]
Pre-selections	9744	8351	14		
$\Delta\phi < 2.0$	9674	8273	14	99	99
$p_T^{\text{lead.}} > 250$ GeV	9625	7461	22	99	90
$N_{b\text{-jet}} < 2$	7640	7461	2	79	100
τ jet veto	7596	7413	2	99	99
$E_T^{\text{miss}} > 600$ GeV	2815	2728	3	37	37
$H_T > 600$ GeV	2815	2728	3	100	100

Table 6.4: Cutflow for the model corresponding to $m_\Phi = 2$ TeV & $R_{\text{inv}} = 0.4$.

The agreement between the ATLAS and MadAnalysis cutflows is globally excellent on both the successive and final number of events, especially concerning the selection on the missing transverse energy that defines the signal region.

I also compared the repartition of the signal events in the 9 bins of the signal region as shown in Tables 6.8 to 6.13. The column «ATLAS» («MA») corresponds to N_{ATLAS} (N_{MA}), the number of weighted signal events in a given bin. The columns «ATLAS proportion» and «MA proportion» account for the number of events in the bin divided by the total number of events in the signal region for ATLAS and MadAnalysis respectively.

	ATLAS	**MA**	**Relative difference [%]**	**ATLAS ϵ [%]**	**MA ϵ [%]**
Pre-selections	10564	10170	4		
$\Delta\phi < 2.0$	10369	10020	3	98	99
$p_T^{\text{lead.}} > 250$ GeV	10284	9500	8	99	95
$N_{b\text{-jet}} < 2$	8694	9500	9	85	100
τ jet veto	8634	9305	8	99	98
$E_T^{\text{miss}} > 600$ GeV	4431	4867	10	51	52
$H_T > 600$ GeV	4430	4867	10	100	100

Table 6.5: Cutflow for the model corresponding to $m_\Phi = 2$ TeV & $R_{\text{inv}} = 0.6$.

	ATLAS	**MA**	**Relative difference [%]**	**ATLAS ϵ [%]**	**MA ϵ [%]**
Pre-selections	298	247	17		
$\Delta\phi < 2.0$	293	238	19	98	97
$p_T^{\text{lead.}} > 250$ GeV	287	214	26	98	90
$N_{b\text{-jet}} < 2$	219	214	2	76	100
τ jet veto	218	212	3	100	99
$E_T^{\text{miss}} > 600$ GeV	59	55	6	27	26
$H_T > 600$ GeV	59	55	6	100	100

Table 6.6: Cutflow for the model corresponding to $m_\Phi = 3$ TeV & $R_{\text{inv}} = 0.2$.

	ATLAS	**MA**	**Relative difference [%]**	**ATLAS ϵ [%]**	**MA ϵ [%]**
Pre-selections	407	360	12		
$\Delta\phi < 2.0$	396	344	13	97	95
$p_T^{\text{lead.}} > 250$ GeV	384	321	16	97	94
$N_{b\text{-jet}} < 2$	315	321	2	82	100
τ jet veto	314	317	1	99	99
$E_T^{\text{miss}} > 600$ GeV	126	128	1	40	40
$H_T > 600$ GeV	126	127	1	100	100

Table 6.7: Cutflow for the model corresponding to $m_\Phi = 3$ TeV & $R_{\text{inv}} = 0.4$.

For the different models, the repartition of the events in the different bins is highly similar between ATLAS and MadAnalysis, demonstrating the great modelization of the two variables p_T^{bal} and $|\phi_{\max} - \phi_{\min}|$.

	ATLAS	**MA**	**ATLAS proportion [%]**	**MA proportion [%]**
Bin 1	3374	3389	3	3
Bin 2	9516	10597	9	10
Bin 3	12812	13214	13	12
Bin 4	15256	16346	15	15
Bin 5	25855	27459	26	26
Bin 6	20496	23168	20	21
Bin 7	4750	5449	5	5
Bin 8	5948	5234	6	5
Bin 9	3227	3647	3	3

Table 6.8: Signal events repartition in the 9 bins of the signal region for the model corresponding to $m_\Phi = 1$ TeV & $R_{\text{inv}} = 0.6$.

	ATLAS	**MA**	**ATLAS proportion [%]**	**MA proportion [%]**
Bin 1	9644	10125	6	6
Bin 2	12579	12957	8	8
Bin 3	11172	10726	7	7
Bin 4	35598	37713	23	23
Bin 5	39655	43161	25	26
Bin 6	24541	24155	15	14
Bin 7	13177	12442	8	8
Bin 8	8609	8409	5	5
Bin 9	4608	4548	3	3

Table 6.9: Signal events repartition in the 9 bins of the signal region for the model corresponding to $m_\Phi = 1$ TeV & $R_{\text{inv}} = 0.8$.

	ATLAS	**MA**	**ATLAS proportion [%]**	**MA proportion [%]**
Bin 1	48	41	2	2
Bin 2	527	450	19	16
Bin 3	609	601	22	22
Bin 4	137	148	5	5
Bin 5	560	576	20	21
Bin 6	540	535	19	20
Bin 7	70	73	2	3
Bin 8	190	168	6	6
Bin 9	134	137	5	5

Table 6.10: Signal events repartition in the 9 bins of the signal region for the model corresponding to $m_\Phi = 2$ TeV & $R_{\text{inv}} = 0.4$.

	ATLAS	**MA**	**ATLAS proportion [%]**	**MA proportion [%]**
Bin 1	135	155	3	3
Bin 2	742	749	17	15
Bin 3	658	758	15	16
Bin 4	382	453	9	9
Bin 5	1040	1144	23	23
Bin 6	733	809	17	17
Bin 7	178	188	4	4
Bin 8	332	367	7	8
Bin 9	229	244	5	5

Table 6.11: Signal events repartition in the 9 bins of the signal region for the model corresponding to $m_\Phi = 2$ TeV & $R_{\text{inv}} = 0.6$.

	ATLAS	**MA**	**ATLAS proportion [%]**	**MA proportion [%]**
Bin 1	0	0	0	0
Bin 2	8	7	13	13
Bin 3	21	21	35	39
Bin 4	1	0	2	0
Bin 5	7	7	12	13
Bin 6	15	13	25	24
Bin 7	1	0	2	0
Bin 8	2	2	3	4
Bin 9	5	4	8	7

Table 6.12: Signal events repartition in the 9 bins of the signal region for the model corresponding to $m_\Phi = 3$ TeV & $R_{\text{inv}} = 0.2$.

	ATLAS	**MA**	**ATLAS proportion [%]**	**MA proportion [%]**
Bin 1	2	2	2	2
Bin 2	17	17	14	14
Bin 3	35	36	28	28
Bin 4	3	4	2	3
Bin 5	19	19	15	15
Bin 6	31	29	24	23
Bin 7	2	2	2	2
Bin 8	6	7	5	6
Bin 9	11	10	8	7

Table 6.13: Signal events repartition in the 9 bins of the signal region for the model corresponding to $m_\Phi = 3$ TeV & $R_{\text{inv}} = 0.4$.

Finally, the observed and expected upper limits on the signal cross-section are computed at 95% confidence level in each of the 9 bins in the SR with the procedure described in section 5.3 in [91], before retaining the limits from the bin that provide the best exclusion. Moreover, it is mentioned in the article that the limits are weaken by about 25% due to systematics uncertainties (mostly from signal modelling). In this way, the limits obtained with MadAnalysis have been multiplied by a factor of 1.25 in order to be compared in a consistent way to the ATLAS limits as shown in Figure 6.7.

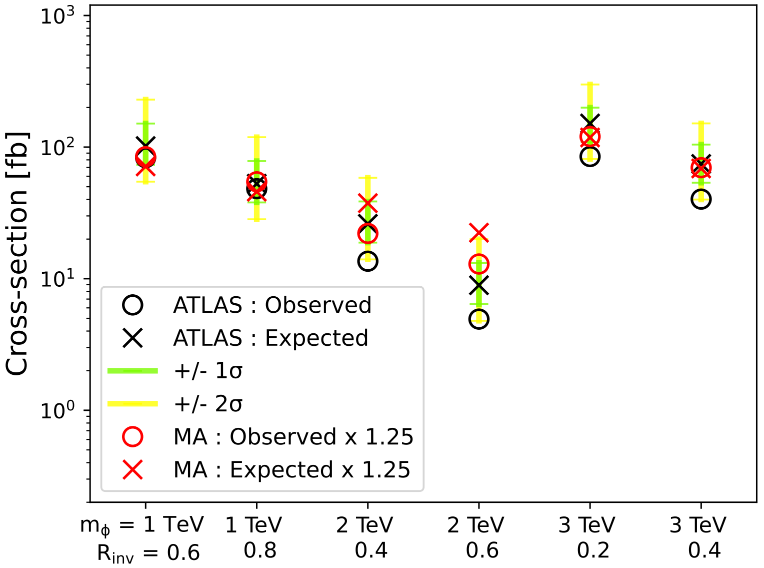


Figure 6.7: Upper limits on the cross-section for the different representative models.

For the 6 representative models, the limits are globally very well reproduced. The analysis has thus been also implemented in the MadAnalysis database and can be reinterpreted with alternative models producing final states with jets in addition to large missing transverse momentum.

6.2.3 ATLAS search for dark jets

Description of the analysis

This analysis was searching for a new Z' boson decaying into a pair of dark quarks (cf Figure 6.8 [97]) which hadronize to form unstable dark mesons π_d and ρ_d that promptly decay back to SM particles. The final state corresponds to events having at least two large-radius jets containing an important number of charged particles. In order to cover a large range of experimental signature, four models have been considered, differing on the number of dark quark flavors, the dark sector confinement scale, the dark particle masses and the dark pion decay modes as shown in Tables 6.14 and 6.15. The common parameters between the four models are the Z' mass, the coupling constant between the Z' and the SM quarks g_q, and the one between the Z' and the dark quarks g_{q_d}.

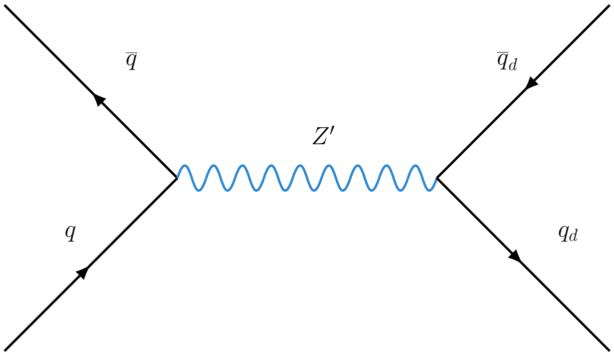


Figure 6.8: Resonant dark quark pair production through a new Z' mediator [97].

Model	n_f	Λ_d (GeV)	$\tilde{m}_{q'}$ (GeV)	m_{π_d} (GeV)	m_{ρ_d} (GeV)	π_d decay mode
A	2	15	20	10	50	$\pi_d \to c\bar{c}$
B	6	2	2	2	4.67	$\pi_d \to s\bar{s}$
C	2	15	20	10	50	$\pi_d \to \gamma'\gamma'$ ($m_{\gamma'} = 4$ GeV)
D	6	2	2	2	4.67	$\pi_d \to \gamma'\gamma'$ ($m_{\gamma'} = 0.7$ GeV)

Table 6.14: Benchmark models used in the analysis : n_f is the number of dark quark flavors, Λ_d is the dark sector confinement scale, $\tilde{m}_{q'}$, m_{π_d} and m_{ρ_d} correspond respectively to the dark quark, dark pion and dark rho masses and γ' represents a dark photon [97]. The decay of the dark rho meson is set to $\rho_d \to \pi_d\pi_d$, while the dark photon decay modes are detailed in Table 6.15.

The analysis uses 139 fb^{-1} of data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector between 2015 and 2018.

The events are first selected with a trigger requiring at least one jet with $p_T > 460$ GeV. The jets of the events are built with topo-clusters that are clustered with the anti-k_t algorithm with a radius of $R = 1.0$. A trimming procedure (cf section 4.4.2) is applied to all the reconstructed jets in order to reduce contributions from pile-up and the underlying event. The jet constituents are

Model C	**Model D**
$\gamma^{'} \rightarrow u\bar{u}$: 22 %	
$\gamma^{'} \rightarrow c\bar{c}$: 22 %	
$\gamma^{'} \rightarrow e^+e^-$: 17 %	$\gamma^{'} \rightarrow \pi^+\pi^-$: 70 %
$\gamma^{'} \rightarrow \mu^+\mu^-$: 17 %	$\gamma^{'} \rightarrow e^+e^-$: 15 %
$\gamma^{'} \rightarrow \tau^+\tau^-$: 10 %	$\gamma^{'} \rightarrow \mu^+\mu^-$: 15 %
$\gamma^{'} \rightarrow d\bar{d}$: 6 %	
$\gamma^{'} \rightarrow s\bar{s}$: 6 %	

Table 6.15: Dark photon decay modes [106].

clustered into sub-jets with the k_t algorithm with $R = 0.2$, and all the sub-jets having their p_T accounting for less than 5% of the p_T of the jet are discarded.

The events must have at least two large-R jets with the (sub-)leading jet satisfying (400) 500 < p_T < 3000 GeV. Moreover, both leading and sub-leading jets must pass $|\eta| < 2.0$ and $50 < m < 600$ GeV, with m the invariant mass of the jet. In addition, a selection on the two leading jets invariant mass $m_{jj} > 1.3$ TeV is applied to ensure that the exploited trigger is fully efficient.

The signal region is then defined with a requirement on the minimum number of tracks associated to the leading and sub-leading jets. The cut value depends on m_{jj}, since requiring a constant minimum value on the jet track multiplicity would affect the shape of the m_{jj} background distribution and impair the background estimation method.

The selection that is applied corresponds to $n_{\text{track},i} > P_{J_i}(m_{jj})$, where $i = 1$ (2) refers to the (sub-)leading jet, and P_{J_i}, shown in Figure 6.9, has been derived in order to remove 99% of the multi-jet background per jet while keeping a sufficient acceptance for the different models. The signal region is thus defined by $n_{\text{track},1} > P_{J_1}(m_{jj})$ & $n_{\text{track},2} > P_{J_2}(m_{jj})$ (or $n^{\epsilon}_{\text{track},i} > 0$ with $n^{\epsilon}_{\text{track},i} = n_{\text{track},i} - P_{J_i}(m_{jj})$).

In this analysis, the m_{jj} spectrum background shape is extracted from data in a control region defined with $n_{\text{track},1} < P_{J_1}(m_{jj})$ & $n_{\text{track},2} < P_{J_2}(m_{jj})$. The m_{jj} background shape is expected to be similar between this control region and the signal region. The background normalization is then determined thanks to a maximum likelihood fit of a given signal-plus-background model to the m_{jj} distribution in data in the signal region.

No significant excess has been observed in data with respect to the expected background contribution. Upper limits on the signal cross-section multiplied by the Z' branching ratio to dark quarks have thus been derived at 95% confidence level using the CLs method with profile likelihood ratio test statistic.

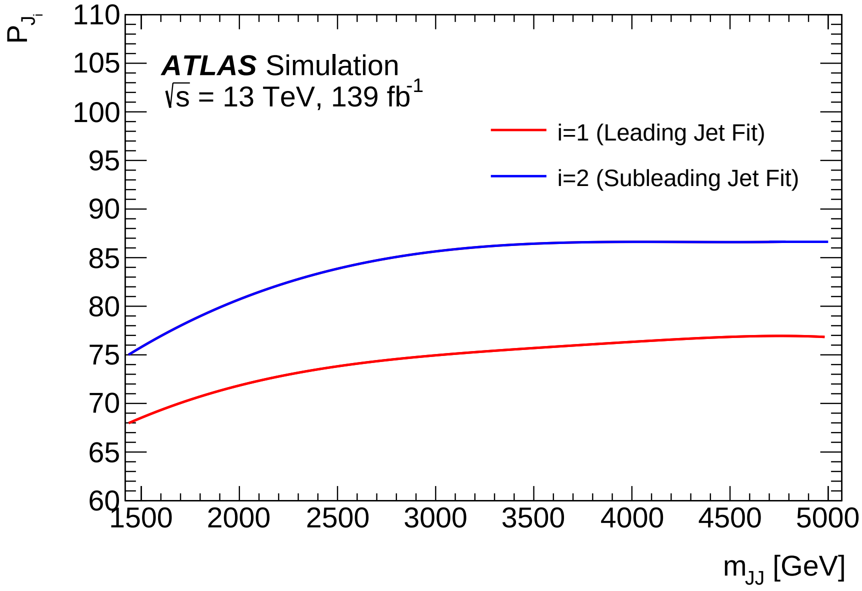


Figure 6.9: Distributions of P_{J_1} and P_{J_2} as a function of m_{jj} for simulated multi-jet events [97].

Implementation and validation

I have written the analysis code that reproduce the selections described previously, including the trimming procedure with the same parameters as ATLAS.

In particular, concerning the selection on the number of tracks for each jet, the parameters of the P_{J_1} and P_{J_2} functions, that correspond to polynomial functions of degree 5 and 3 respectively, are presented in Table 6.16, and have been taken from ATLAS fit results. For the leading and sub-leading jets, the number of tracks $n_{\mathrm{track},i}$ has been determined by counting the charged and visible particles in the final state that are within $\Delta R < 1.2$ of the jet axis.

	P_{J_1}	P_{J_2}
p_0	$4.65.10^{1}$	$4.66.10^{1}$
p_1	$2.12.10^{-2}$	$2.82.10^{-2}$
p_2	$-3.91.10^{-6}$	$-6.61.10^{-6}$
p_3	$-6.98.10^{-10}$	$5.15.10^{-10}$
p_4	$3.22.10^{-13}$	/
p_5	$-2.94.10^{-17}$	/

Table 6.16: Parameters of the polynomial function $P_{J_i}(X) = \sum_k p_k X^k$.

Moreover, a smearing of the energy and angular coordinates η and ϕ is applied to the jet constituents instead of the global jet as it is usually done. The constituent energy is smeared through the multiplication of the initial energy by a random number generated accordingly to a Gaussian distribution with a mean of 1 and a width of 0.1. Concerning the angular coordinates, the initial values are summed to a random number generated accordingly to a Gaussian distribution having a mean of 0 and a width of $\sigma(p_T) = \alpha/(1 + e^{\beta(p_T - \gamma)})$ [107], with the parameter values set to $\alpha = 0.1$, $\beta = 0.02$ GeV^{-1} and $\gamma = 2$ GeV. These values have been obtained by iteration in order to improve the agreement with ATLAS concerning the efficiency of the cut on the jet mass.

Since ATLAS provides cutflows for the $m_{Z'} = 2.5$ TeV signal for the four different models, I have generated around 90k events for each of them with Pythia 8.308 [78] using the same parameters.

The cutflows are presented in Tables 6.17 to 6.20. In the «ATLAS» («MadAnalysis») column, the numbers correspond to the relative efficiency in % of a given selection, while the last line represents the global acceptance. The agreement between ATLAS and MadAnalysis is globally very good.

Model A	**ATLAS**	**MadAnalysis**
Trigger, $m_{jj} > 1.3$ TeV	80	85
$m_{j1,2} > 50$ GeV, $p_{T,j1} > 500$ GeV, $p_{T,j2} > 400$ GeV	88	90
$\|\eta_{j1,2}\| < 2$	100	97
$m_{j1,2} < 600$ GeV, $p_{T,j1,2} < 3000$ GeV	100	100
$n^{\epsilon}_{\text{track},j1,2} > 0$	37	38
Total	26	28

Table 6.17: Cutflow for the signal from the model A corresponding to $m_{Z'} = 2.5$ TeV. The numbers correspond to the relative efficiency in % of a given selection, while the last line represents the global acceptance.

Model B	**ATLAS**	**MadAnalysis**
Trigger, $m_{jj} > 1.3$ TeV	84	88
$m_{j1,2} > 50$ GeV, $p_{T,j1} > 500$ GeV, $p_{T,j2} > 400$ GeV	60	60
$\|\eta_{j1,2}\| < 2$	100	98
$m_{j1,2} < 600$ GeV, $p_{T,j1,2} < 3000$ GeV	100	100
$n^{\epsilon}_{\text{track},j1,2} > 0$	3	2
Total	1	1

Table 6.18: Cutflow for the signal from the model B corresponding to $m_{Z'} = 2.5$ TeV. The numbers correspond to the relative efficiency in % of a given selection, while the last line represents the global acceptance.

In order to derive the cross-section upper limits, the background normalization has to be determined for a given signal. To compute it, a fit is performed by minimizing the function :

$$-2\log(L(\mu,\ \nu(\mu))) = -2\sum_i (n_i \log(\nu(\mu)b_i + \mu s_i) - \nu(\mu)b_i - \mu s_i) \tag{6.1}$$

where the sum is done on the m_{jj} bins, n_i and b_i correspond respectively to the observed number of events and the nominal values for the expected background in the bin i that have been both published by ATLAS, s_i is the number of signal events obtained with MadAnalysis, ν is the background

Model C	**ATLAS**	**MadAnalysis**
Trigger, $m_{jj} > 1.3$ TeV	66	67
$m_{j1,2} > 50$ GeV, $p_{T,j1} > 500$ GeV, $p_{T,j2} > 400$ GeV	81	80
$\lvert\eta_{j1,2}\rvert < 2$	100	98
$m_{j1,2} < 600$ GeV, $p_{T,j1,2} < 3000$ GeV	100	100
$n^{\epsilon}_{\mathrm{track},j1,2} > 0$	12	12
Total	6	6

Table 6.19: Cutflow for the signal from the model C corresponding to $m_{Z'} = 2.5$ TeV. The numbers correspond to the relative efficiency in % of a given selection, while the last line represents the global acceptance.

Model D	**ATLAS**	**MadAnalysis**
Trigger, $m_{jj} > 1.3$ TeV	79	81
$m_{j1,2} > 50$ GeV, $p_{T,j1} > 500$ GeV, $p_{T,j2} > 400$ GeV	56	51
$\lvert\eta_{j1,2}\rvert < 2$	100	98
$m_{j1,2} < 600$ GeV, $p_{T,j1,2} < 3000$ GeV	100	100
$n^{\epsilon}_{\mathrm{track},j1,2} > 0$	55	65
Total	24	26

Table 6.20: Cutflow for the signal from the model D corresponding to $m_{Z'} = 2.5$ TeV. The numbers correspond to the relative efficiency in % of a given selection, while the last line represents the global acceptance.

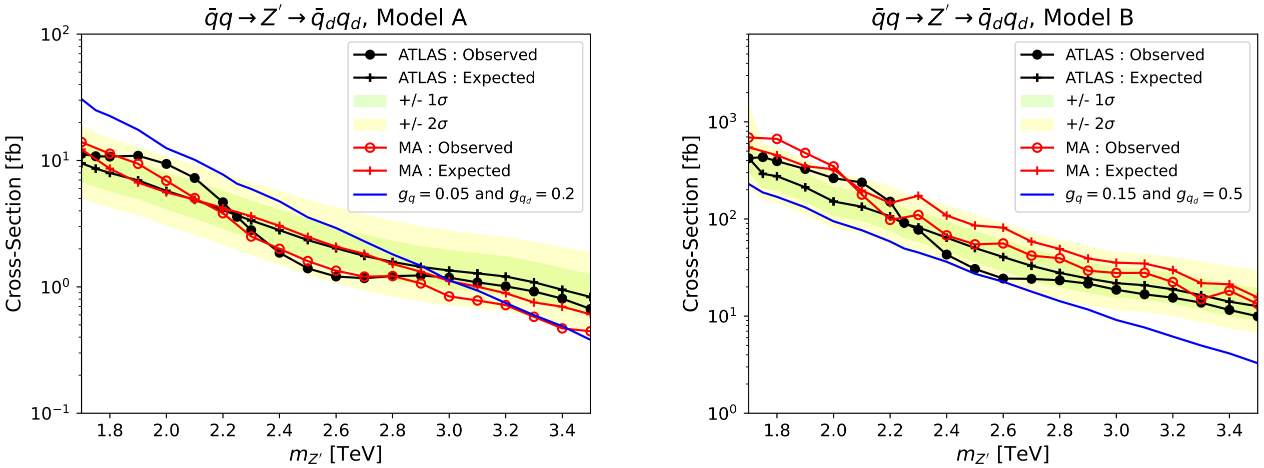
normalization and μ is the signal strength. The best fit value $\hat{\mu}$ that minimizes this function is then used to determine the corresponding normalization $\nu(\hat{\mu})$. The comparison of the normalizations obtained with MadAnalysis for the different signals with the ones obtained by ATLAS (that have been get from the ATLAS internal note) shows an excellent agreement.

Then, multiple sub-regions of the signal region are created according to the m_{jj} binning of 100 GeV used by ATLAS : a sub-region with $m_{jj} > 1.3$ TeV, one with $m_{jj} > 1.4$ TeV, etc, until a region with $m_{jj} > 4.9$ TeV (since ATLAS investigated m_{jj} values between 1.3 and 5 TeV). In each of these sub-regions, the number of observed events and the post-fit number of background events are computed. The background uncertainty on a given number of background b is taken as $\sqrt{b + (b/10)^2}$, such as the quadratic sum of a statistical error and a systematic uncertainty of 10% (which is consistent with ATLAS results).

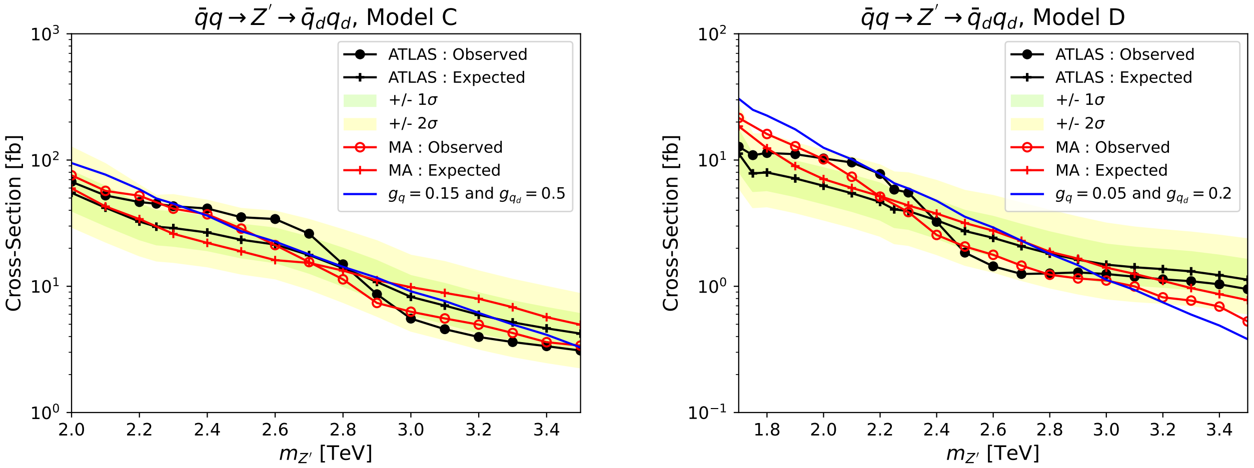
Finally, the upper limits on the signal cross-section are computed in each of these sub-regions with the CLs method using the same procedure as the previous analyses, and the limit corresponding to the best exclusion is retained.

Concerning the upper limits, ATLAS provided the values for several masses between 1.5 and 3.5 TeV for model A, B and D, and between 2 and 3.5 TeV for model C. For each mass, 20k events have been generated.

The observed and expected upper limits on the cross-section at 95% confidence level for ATLAS and MadAnalysis are presented as a function of $m_{Z'}$ for the four different models in Figure 6.10, and the agreement is found to be good. This analysis has thus been validated and will be soon added to the MadAnalysis public database.



(a) Model A (b) Model B



(c) Model C (d) Model D

Figure 6.10: Upper limits on the cross-section for the four different models.

6.3 Reinterpretation of the ATLAS Run-3 emerging jet analysis

Concerning the emerging jet analysis presented in section 5, it is also important to publish and provide the relevant information that can allow future reinterpretation of the analysis (especially the cut-based one) with alternative models producing emerging jets.

For example, to improve the accuracy of the reinterpretation of the analysis, the statistical model (i.e the likelihood function) has to be provided in a format that can be exploited by phenomenologists that do not belong to the ATLAS collaboration. Indeed, for each of the analyses presented in section 6.2, I had to implement a method to interpret the results that was different from the one employed in the analysis since the collaboration did not provide the corresponding statistical model.

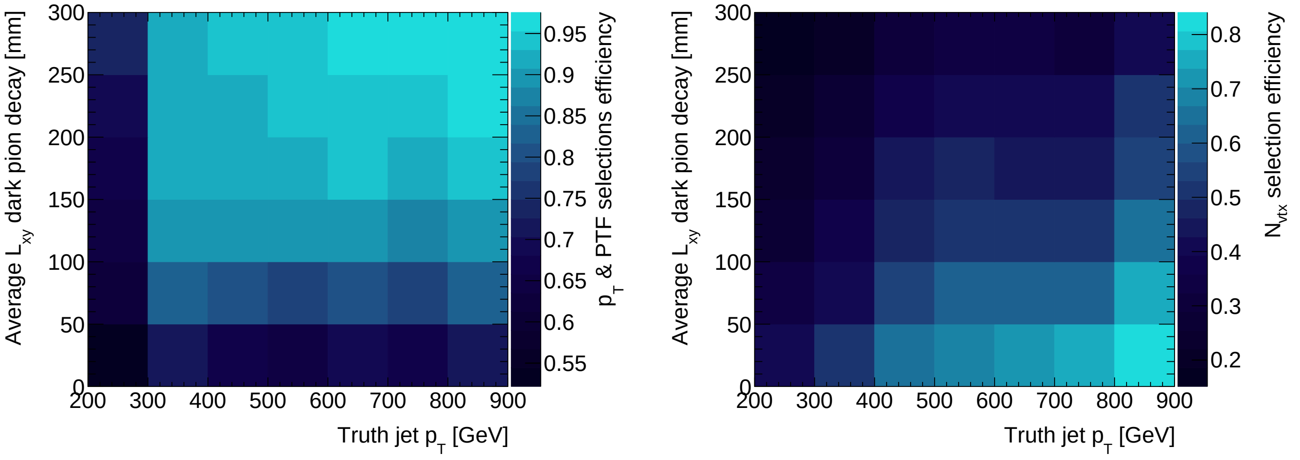
Moreover, a way to allow future reinterpretation of the cut-based analysis could be to provide the efficiencies of the selections applied on variables that are difficult to reproduce externally using reinterpretation frameworks. These efficiencies could be parametrized as a function of truth information. Concerning the low-m_{jj} cut-based analysis, the two selections that are concerned are the ones on PTF and N_{vtx}, since these variables are based on a full simulation of the inner detector and on the tracking and vertexing ATLAS algorithms, while the other selections are easily reproducible.

To derive these efficiencies, I considered signal events corresponding to $m_{Z'} \in \{0.6,\ 1.5,\ 3\}$ TeV, $m_{\pi_d} \in \{5,\ 10,\ 20\}$ GeV and $c\tau_{\pi_d} \in \{5,\ 50\}$ mm. As a first attempt, the efficiencies have been evaluated in bins of (p_T^{truth}, $<\!L_{xy}\!>$), where p_T^{truth} corresponds to the p_T of the truth jet (the trigger-matched one) and $<\!L_{xy}\!>$ is equal to $\sum(L_{xy}^{\text{vtx}} \times p_T^{\text{vtx}})/\sum p_T^{\text{vtx}}$ with the sums occurring on the dark pion decay vertices that are geometrically associated to the trigger-matched jet, L_{xy}^{vtx} is the distance in the transverse plane between the primary vertex and the dark pion decay vertex and p_T^{vtx} is the transverse momentum associated to the vertex. These variables, and especially $<\!L_{xy}\!>$, are expected to be correlated to the efficiency of selections on PTF and N_{vtx}. The efficiencies can then be visualized as 2D maps with these two variables defining the X and Y axis.

The first efficiency map corresponds to the selections «Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04», with the selections on p_T and PTF applied to the trigger-matched jet. The signal events are required to satisfy $N_{\text{jet}} \geq 2$. Then, the efficiencies are evaluated in bins of (p_T^{truth}, $<\!L_{xy}\!>$) as the proportion of events passing the selections. The corresponding efficiency map is shown in Figure 6.11.a : it is possible to notice that the efficiencies are consistently increasing with $<\!L_{xy}\!>$, since high values of $<\!L_{xy}\!>$ are likely to correspond to low PTF values.

The second efficiency map corresponds to the selection «$N_{\text{vtx}} \geq 1$» applied to the leading and sub-leading jets. The signal events are required to satisfy the selections «Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04 & $|\eta^{\text{lead.}}| < 1.5$ & $|\eta^{\text{sub-lead.}}| < 1.5$», and the efficiencies then correspond to the proportion of events passing the selection on N_{vtx}. The corresponding map is shown in Figure 6.11.b : concerning this selection, the efficiencies are higher at lower $<\!L_{xy}\!>$ values since the vertex reconstruction efficiency is likely to decrease with $<\!L_{xy}\!>$.

By evaluating the statistical uncertainty in the efficiency using $\sigma(\epsilon_i) = \sqrt{\epsilon_i(1+\epsilon_i)}/\sqrt{n_{\text{jet},i}}$, with ϵ_i and $n_{\text{jet},i}$ the efficiency and the number of jets in the bin i, the relative uncertainty in the efficiency is at maximum equal to 1.5 (3) % for the map corresponding to the selection on PTF (N_{vtx}).



(a) (b)

Figure 6.11: Efficiency map for the selections corresponding to (a) «Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04» and (b) «$N_{\text{vtx}} \geq 1$». The bins at the border of the maps include the overflow.

To evaluate at which level this parametrization is sufficiently accurate to reproduce the analysis cutflow, a test has been performed : it consists in computing the number of signal events passing the several selections by considering the computed efficiencies for the selections on PTF and N_{vtx}, and to compare them to the numbers computed by the simple application of the selections. More precisely, for a given signal, the selection $N_{\text{jet}} \geq 2$ is firstly applied to the events. Then, the per-event efficiency for the selections «Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04» is obtained from the corresponding map, and the event weight is multiplied by this efficiency. The events are then required to pass the selections «$|\eta^{\text{lead.}}| < 1.5$ & $|\eta^{\text{sub-lead.}}| < 1.5$». The event weight is then re-multiplied by the corresponding efficiency concerning the selection «$N_{\text{vtx}} \geq 1$» obtained from the dedicated map. By summing the event weights at each step of the cutflow, it is thus possible to obtain a cutflow that exploits the efficiency maps for the selections on PTF and N_{vtx}.

These cutflows are presented in Tables 6.21 to 6.24 for four different models. The numbers in the column «Cuts» correspond to the numbers of signal events obtained with the standard application of the selections, while the ones in «Maps» correspond to the numbers obtained by using the two efficiency maps. Non negligible disagreements between these two cutflows can be noticed, even with finer map binnings. It means that the parametrization of the efficiencies using only two variables is not sufficiently precise.

An alternative method that is currently investigated would be to use a dedicated neural network taking as inputs several truth information in order to predict for a given signal the efficiencies of the selections on PTF and N_{vtx}. That would permit to consider a larger number of variables and their correlations to parametrize the efficiencies and to potentially obtain a greater accuracy for reinterpretation purposes.

	Cuts	**Maps**	Relative difference [%]
$N_{\text{jet}} \geq 2$	26848	26848	0
Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04	15244	17257	13
Leading & sub-leading jets : $\lvert\eta\rvert < 1.5$	10888	12741	17
Leading & sub-leading jets : $N_{\text{vtx}} \geq 1$	3257	4733	45

Table 6.21: Cutflows comparison between the standard cutflow («Cuts») and the one obtained using the efficiency maps («Maps») for the signal corresponding to $m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 5$ mm and $m_{Z'} = 0.6$ TeV. The relative difference corresponds to $|N_{\text{cuts}} - N_{\text{maps}}|/N_{\text{cuts}}$ with N the number of events at a given step.

	Cuts	**Maps**	Relative difference [%]
$N_{\text{jet}} \geq 2$	29534	29534	0
Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04	21401	21593	1
Leading & sub-leading jets : $\lvert\eta\rvert < 1.5$	15830	16158	2
Leading & sub-leading jets : $N_{\text{vtx}} \geq 1$	1771	4069	130

Table 6.22: Cutflows comparison between the standard cutflow («Cuts») and the one obtained using the efficiency maps («Maps») for the signal corresponding to $m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 50$ mm and $m_{Z'} = 0.6$ TeV.

	Cuts	**Maps**	Relative difference [%]
$N_{\text{jet}} \geq 2$	80732	80732	0
Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04	59566	65347	10
Leading & sub-leading jets : $\lvert\eta\rvert < 1.5$	44418	48921	10
Leading & sub-leading jets : $N_{\text{vtx}} \geq 1$	22646	26366	16

Table 6.23: Cutflows comparison between the standard cutflow («Cuts») and the one obtained using the efficiency maps («Maps») for the signal corresponding to $m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 5$ mm and $m_{Z'} = 1.5$ TeV.

	Cuts	**Maps**	Relative difference [%]
$N_{\text{jet}} \geq 2$	84226	84226	0
Emerging jet trigger & $p_T > 250$ GeV & PTF < 0.04	66665	72132	8
Leading & sub-leading jets : $\lvert\eta\rvert < 1.5$	53779	57942	8
Leading & sub-leading jets : $N_{\text{vtx}} \geq 1$	31743	35787	13

Table 6.24: Cutflows comparison between the standard cutflow («Cuts») and the one obtained using the efficiency maps («Maps») for the signal corresponding to $m_{\pi_d} = 5$ GeV, $c\tau_{\pi_d} = 5$ mm and $m_{Z'} = 3$ TeV.

6.4 Conclusion

This section presented the principle of the reinterpretation of BSM searches performed at the LHC. It also presented my contribution concerning the implementation of several ATLAS analyses within the MadAnalysis public database. Studies that exploit these analyses in addition to other ones that are sensitive to dark QCD models are currently in finalization within a small international collaboration of about 10 experimentalists and theorists in order to constrain the global dark QCD parameter phase space.

Moreover, allowing the reinterpretation of the emerging jet analysis by providing useful information together with the article is fundamental. I am currently involved in the evaluation of the selection efficiencies concerning the variables used in the analysis that are difficult to reproduce with reinterpretation frameworks.

Conclusion

This document presented a search for emerging jets with the ATLAS detector considering an s-channel Z' mediator that acts as a portal between a QCD-like dark sector and the visible sector. During my thesis, I contributed in a significant way at different stages to one of the channels of selection of the analysis. The final results demonstrates a remarkable improvement concerning the sensitivity to new heavy Z' mediator. In the future, this analysis will be re-done with different strategies and exploit the full set of Run-3 data that corresponds to an integrated luminosity of about 350 fb^{-1}. It is thus expected to lead to an important gain in sensitivity since the analysis results presented in this thesis are limited by the available statistics. An other limiting point in the analysis is the degradation in terms of track and vertex reconstruction efficiency at higher radius in the inner detector, while the investigated final state corresponds to jets associated to displaced objects. The replacement of the current inner detector with «ITk», a full silicon-technology detector aiming to reconstruct tracks with a better efficiency within the high-luminosity LHC conditions, is thus expected to increase significantly the sensitivity to the emerging jet topology.

This thesis also presented the reinterpretation of LHC analyses, that aims to re-employ existing results from ATLAS or CMS and to constrain alternative models. I notably contributed to the implementation within the MadAnalysis public database of several analyses that are sensitive to dark QCD models, with the aim of constraining the corresponding parameter phase space. At the time this thesis is written, I am still investigating possibilities to allow the future reinterpretation of the emerging jet analysis by providing information that are exploitable and pertinent for such purpose.

Finally, during my thesis, I also became an author of the ATLAS collaboration by proposing modifications in the algorithms employed in ATLAS to reconstruct the jet constituents, allowing notably to improve the reconstruction performances for jets located at the edge of the inner detector acceptance.

Bibliography

[1] Peter W. Higgs. *Broken Symmetries and the Masses of Gauge Bosons*. Phys. Rev. Lett. 13 (1964), p.508-509. DOI : 10.1103/PhysRevLett.13.508.

[2] David Galbraith and Carsten Burgard. *Standard model of physics*. (2012). URL : `https://texample.net/model-physics/`

[3] The CDF Collaboration. *Observation of Top Quark Production in $\bar{p}p$ Collisions*. Physical Review Letters 74 (1995), p.2626–2631. DOI : 10.1103/physrevlett.74.2626. arXiv:hep-ex/9503002

[4] The D0 Collaboration. *Observation of the Top Quark*. Physical Review Letters 74 (1995), p.2632–2637. DOI : 10.1103/physrevlett.74.2632. arXiv:hep-ex/9503003

[5] The DONUT Collaboration. *Observation of tau neutrino interactions*. Physics Letters B 504 (2001), p.218–224. DOI : 10.1016/s0370-2693(01)00307-0. arXiv:hep-ex/0012035

[6] The ATLAS Collaboration. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector*. Physics Letters B 716 (2012), p.1–29. DOI : 10.1016/j.physletb.2012.08.020. arXiv:1207.7214

[7] The CMS Collaboration. *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. Physics Letters B 716 (2012), p.30–61. DOI : 10.1016/j.physletb.2012.08.021. arXiv:1207.7235

[8] Wikipedia. *Feynman diagram*. (2017). URL : `https://fr.wikipedia.org/wiki/Diagramme_de_Feynman`

[9] Particle Data Group. *Review of Particle Physics*. Progress of Theoretical and Experimental Physics 2020, 083C01 (2020). DOI : 10.1093/ptep/ptaa104.

[10] Stefan Höche. *Introduction to parton-shower event generators*. (2015). arXiv:1411.4085

[11] Fritz Zwicky. *The Redshift of Extragalactic Nebulae*. Helvetica Physica Acta 6 (1933), p.110-127.

[12] Rubin, V. C. & Ford, W. K., Jr. *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*. Astrophysical Journal 159 (1970), p.379

[13] McGaugh, Stacy. *The Third Law of Galactic Rotation*. Galaxies 2(4) (2014), p.601. DOI : 10.3390/galaxies2040601. arXiv:1412.3767

[14] The Planck Collaboration. *Planck 2018 results: VI. Cosmological parameters*. Astronomy & Astrophysics 641 (2020), A6. DOI : 10.1051/0004-6361/201833910. arXiv:1807.06209

[15] The ATLAS Collaboration. *Searching for Dark Matter with the ATLAS detector*. (2019). URL : `https://atlas.cern/updates/feature/dark-matter`

[16] Arbey, A. and Mahmoudi, F. *Dark matter and the early Universe: a review*. Prog. Part. Nucl. Phys. 119 (2021), p.103865. DOI : 10.1016/j.ppnp.2021.103865. arXiv:2104.11488

[17] The ATLAS Collaboration. *Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at* $\sqrt{s}$ *=13 TeV with the ATLAS detector*. Physical Review D 103 (2021), p.112006. DOI : 10.1103/physrevd.103.112006. arXiv:2102.10874

[18] Park, Myeonghun and Zhang, Mengchao. *Tagging a jet from a dark sector with jet substructures at colliders*. Physical Review D 100 (2019), p.115009. DOI : 10.1103/physrevd.100.115009. arXiv:1712.09279

[19] E. Lopienska. *The CERN accelerator complex, layout in 2022*. (2022). URL : `https://cds.cern.ch/record/2800984?ln=fr`

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

[21] The CMS Collaboration. *The CMS experiment at the CERN LHC*. JINST 3 (2008) S08004. DOI : 10.1088/1748-0221/3/08/S08004.

[22] The ALICE Collaboration. *The ALICE experiment at the CERN LHC*. JINST 3 (2008) S08002. DOI : 10.1088/1748-0221/3/08/S08002.

[23] The LHCb Collaboration. *The LHCb Detector at the LHC*. JINST 3 (2008) S08005. DOI : 10.1088/1748-0221/3/08/S08005.

[24] The ATLAS Collaboration. *The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3*. Journal of Instrumentation 19 (2024). DOI : 10.1088/1748-0221/19/05/p05063. arXiv:2305.16623

[25] *CMS coordinate system*. URL : `https://tikz.net/axis3d_cms/`

[26] Quantum Diaries. *ATLAS magnet system*. URL : `https://www.quantumdiaries.org/wp-content/uploads/2011/05/exp-magnets.png`

[27] The ATLAS Collaboration. *ATLAS Detector Magnet System*. URL : `https://cds.cern.ch/images/ATLAS-PHOTO-2021-029-1`

[28] W. Herr and B. Muratori. *Concept of luminosity*. URL : `https://cds.cern.ch/record/941318/files/p361.pdf`

[29] The ATLAS Collaboration. *Public ATLAS Luminosity Results for Run-3 of the LHC*. URL : `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3`

[30] Sjöstrand, Torbjörn et al. *An introduction to PYTHIA 8.2*. Computer Physics Communications 191 (2015), p.159. DOI : 10.1016/j.cpc.2015.01.024. arXiv:1410.3012

[31] E. Bothmann et al. *Event generation with Sherpa 2.2*. SciPost Phys. 7 (2019), p.034. arXiv:1905.09127

[32] G. Bewick et al. *Herwig 7.3 Release Note*. (2024). arXiv:2312.05175

[33] Alwall, J. et al. *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*. Journal of High Energy Physics 07 (2014), p.079. DOI : 10.1007/jhep07(2014)079. arXiv:1405.0301v2

[34] The GEANT4 Collaboration. *GEANT4–a simulation toolkikt*. Nucl. Instrum. Methods Phys. Res., A 506 (2003), p.250-303. DOI : 10.1016/S0168-9002(03)01368-8.

[35] The ATLAS Collaboration. *ATLAS Track Reconstruction*. (2024). URL : `https://atlassoftwaredocs.web.cern.ch/internal-links/tracking-tutorial/idoverview/`.

[36] The ATLAS Collaboration. *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*. Computing and Software for Big Science 8 (2024), p.9. DOI : 10.1007/s41781-023-00111-y. arXiv:2308.09471

[37] R. Frühwirth. *Application of Kalman filtering to track and vertex fitting*. Nuclear Instruments and Methods in Physics Research Section A 262 (1987), p444-450. DOI : 10.1016/0168-9002(87)90887-4.

[38] The ATLAS Collaboration. *Performance of the reconstruction of large impact parameter tracks in the inner detector of ATLAS*. The European Physical Journal C 83 (2023), p.1081. DOI : 10.1140/epjc/s10052-023-12024-6. arXiv:2304.12867

[39] The ATLAS Collaboration. *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*. (2019). URL : `https://cds.cern.ch/record/2670380/files/ATL-PHYS-PUB-2019-015.pdf`.

[40] The ATLAS Collaboration. *Performance of vertex reconstruction algorithms for detection of new long-lived particle decays within the ATLAS inner detector*. (2019). URL : `https://cds.cern.ch/record/2669425/files/ATL-PHYS-PUB-2019-013.pdf`.

[41] The ATLAS Collaboration. *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*. The European Physical Journal C 77 (2017), p.490. DOI : 10.1140/epjc/s10052-017-5004-5. arXiv:1603.02934

[42] The ATLAS Collaboration. *Jet reconstruction and performance using particle flow with the ATLAS Detector*. The European Physical Journal C 77.7 (2017), p.466. DOI : 10.1140/epjc/s10052-017-5031-2. arXiv:1703.10485

[43] The ATLAS Collaboration. *Improving jet substructure performance in ATLAS using Track-CaloClusters*. (2017). URL : `https://cds.cern.ch/record/2275636`.

[44] The ATLAS Collaboration. *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions*. The European Physical Journal C 81.4 (2021), p.334. DOI : 10.1140/epjc/s10052-021-09054-3. arXiv:2009.04986

[45] Berta, Peter and Spousta, Martin and Miller, David W. and Leitner, Rupert. *Particle-level pileup subtraction for jets and jet shapes*. Journal of High Energy Physics 06 (2014), p.92. DOI : 10.1007/jhep06(2014)092. arXiv:1403.3108

[46] The ATLAS Collaboration. *Constituent-level pile-up mitigation techniques in ATLAS*. (2017). URL : `https://cds.cern.ch/record/2281055/?ln=fr`

[47] Cacciari, Matteo and Salam, Gavin P. and Soyez, Gregory. *SoftKiller, a particle-level pileup removal method*. The European Physical Journal C 75 (2015). DOI : 10.1140/epjc/s10052-015-3267-2. arXiv:1407.0408

[48] Cacciari, Matteo and Salam, Gavin P and Soyez, Gregory. *The anti-k_t jet clustering algorithm*. Journal of High Energy Physics 04 (2008), p.63. DOI : 10.1088/1126-6708/2008/04/063. arXiv:0802.1189

[49] Cacciari, Matteo and Salam, Gavin P. and Soyez, Gregory. *FastJet user manual*. The European Physical Journal C 72 (2012), p.1896. DOI : 10.1140/epjc/s10052-012-1896-2. arXiv:1111.6097

[50] Krohn, David and Thaler, Jesse and Wang, Lian-Tao. *Jet trimming*. Journal of High Energy Physics 02 (2010), p.84. DOI : 10.1007/jhep02(2010)084. arXiv:0912.1342

[51] Larkoski, Andrew J. and Marzani, Simone and Soyez, Gregory and Thaler, Jesse. *Soft drop*. Journal of High Energy Physics 05 (2014), p.146. DOI : 10.1007/jhep05(2014)146. arXiv:1402.2657

[52] The ATLAS Collaboration. *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. The European Physical Journal C 81 (2021), p.689. DOI : 10.1140/epjc/s10052-021-09402-3. arXiv:2007.02645

[53] The ATLAS Collaboration. *R=0.4 jets input comparison and Monte Carlo calibration with the ATLAS Detector*. (2022). URL : `https://cds.cern.ch/record/2824558/?ln=fr`.

[54] Ball, Richard D. et al. *Parton distributions with LHC data*. Nuclear Physics B 867 (2013), p.244. DOI : 10.1016/j.nuclphysb.2012.10.003. arXiv:1207.1303

[55] The ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. (2014). URL : `https://cds.cern.ch/record/1966419?ln=fr`.

[56] The ATLAS Collaboration. *Measurement of jet-substructure observables in top quark, W boson and light jet production in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Journal of High Energy Physics 08 (2019), p.033. DOI : 10.1007/jhep08(2019)033. arXiv:1903.02942

[57] The ATLAS Collaboration. *Search for emerging jets in pp collisions at $\sqrt{s} = 13.6$ TeV with the ATLAS experiment*. (2025). arXiv:2505.02429

[58] Schwaller, Pedro and Stolarski, Daniel and Weiler, Andreas. *Emerging jets*. Journal of High Energy Physics 05 (2015), p.59. DOI : 10.1007/jhep05(2015)059. arXiv:1502.05409

[59] Cohen, Timothy and Lisanti, Mariangela and Lou, Hou Keong. *Semivisible Jets: Dark Matter Undercover at the LHC*. Physical Review Letters 115 (2015), p.171804. DOI : 10.1103/physrevlett.115.171804. arXiv:1503.00009

[60] The ATLAS Collaboration. *ATLAS data quality operations and performance for 2015–2018 data-taking*. Journal of Instrumentation 15 (2020). DOI : 10.1088/1748-0221/15/04/p04003. arXiv:1911.04632

[61] Stefano Frixione, Paolo Nason, Giovanni Ridolfi. *A Positive-Weight Next-to-Leading-Order Monte Carlo for Heavy Flavour Hadroproduction*. Journal of High Energy Physics 09 (2007), p.126. DOI : 10.1088/1126-6708/2007/09/126. arXiv:0707.3088

[62] Paolo Nason. *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*. Journal of High Energy Physics 11 (2004), p.040. DOI : 10.1088/1126-6708/2004/11/040. arXiv:hep-ph/0409146

[63] Stefano Frixione, Paolo Nason, Carlo Oleari. *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*. Journal of High Energy Physics 11 (2007), p.070. DOI : 10.1088/1126-6708/2007/11/070. arXiv:0709.2092

[64] Simone Alioli, Paolo Nason, Carlo Oleari, Emanuele Re. *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*. Journal of High Energy Physics 06 (2010), p.043. DOI : 10.1007/JHEP06/282010/29043. arXiv:1002.2581

[65] The NNPDF Collaboration. *Parton distributions for the LHC Run II*. Journal of High Energy Physics 04 (2015), p.040. DOI : 10.1007/JHEP04/282015/29040. arXiv:1410.8849

[66] D.J. Lange. *The EvtGen particle decay simulation package*. Nuclear Instruments and Methods in Physics Research Section A 462 (2001), p.152. DOI : 10.1016/S0168-9002(01)00089-4.

[67] T. Gleisberg and S. Höche. *Comix, a new matrix element generator*. Journal of High Energy Physics 12 (2008), p.039. arXiv:0808.3674

[68] F. Buccioni et al. *OpenLoops 2*. Eur. Phys. J. C 79 (2019), p.866. arXiv:1907.13071

[69] F. Cascioli, P. Maierhöfer and S. Pozzorini. *Scattering Amplitudes with Open Loops*. Phys. Rev. Lett. 108 (2012), p.111601. arXiv:1111.5206

[70] A. Denner, S. Dittmaier and L. Hofer. *Collier: A fortran-based complex one-loop library in extended regularizations*. Comput. Phys. Commun. 212 (2017), p.220. arXiv:1604.06792

[71] S. Schumann and F. Krauss. *A parton shower algorithm based on Catani–Seymour dipole factorisation*. Journal of High Energy Physics 03 (2008), p.038. arXiv:0709.1027

[72] S. Höche, F. Krauss, M. Schönherr and F. Siegert. *A critical appraisal of NLO+PS matching methods*. Journal of High Energy Physics 09 (2012), p.049. arXiv:1111.1220

[73] S. Höche, F. Krauss, M. Schönherr and F. Siegert. *QCD matrix elements + parton showers. The NLO case*. Journal of High Energy Physics 04 (2013), p.027. arXiv:1207.5030

[74] S. Catani, F. Krauss, B. R. Webber and R. Kuhn. *QCD Matrix Elements + Parton Showers*. Journal of High Energy Physics 11 (2002), p.063. arXiv:hep-ph/0109231

[75] S. Höche, F. Krauss, S. Schumann and F. Siegert. *QCD matrix elements and truncated showers*. Journal of High Energy Physics 05 (2009), p.053. arXiv:0903.1219

[76] L. Carloni and T. Sjöstrand. *Visible effects of invisible hidden valley radiation*. Journal of High Energy Physics 09 (2010), p.105 arXiv:1006.2911

[77] L. Carloni, J. Rathsman and T. Sjöstrand. *Discerning secluded sector gauge structures*. Journal of High Energy Physics 04 (2011), p.091. arXiv:1102.3795

[78] C. Bierlich et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. SciPost Phys. Codeb. (2022), p.8. arXiv:2203.11601

[79] K. Werner, F.-M. Liu and T. Pierog. *Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at the BNL Relativistic Heavy Ion Collider*. Phys. Rev. C 74 (2006), p.044902. arXiv:hep-ph/0506232

[80] W.Buttinger. *Background Estimation with the ABCD Method*. URL : `https://www.semanticscholar.org/paper/Background-Estimation-with-the-ABCD-Method-the-Buttinger/ab7cef9e8e5079ad7083a5369a0e9c87aada56fa`

[81] Cowan, G. et al. *Asymptotic formulae for likelihood-based tests of new physics*. The European Physical Journal C 71 (2011), p.1554. DOI : 10.1140/epjc/s10052-011-1554-0. arXiv:1007.1727

[82] A. L. Read, *Presentation of search results: the CL_s technique*. J. Phys. G: Nucl. Part. Phys. 28 (2002), p.2693-2704.

[83] L. Heinrich, M. Feickert and G. Stark. *scikit-hep/pyhf: v0.7.6*. URL : `https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6`

[84] L. Heinrich, M. Feickert, G. Stark and K. Cranmer. *pyhf: pure-Python implementation of HistFactory statistical models*. J. Open Source Softw. 6 (2021), p.2823

[85] The ATLAS Collaboration. *Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*. Journal of High Energy Physics 03 (2020), p.145. DOI : 10.1007/jhep03(2020)145. arXiv:1910.08447

[86] The ATLAS Collaboration. *Constraints on simplified dark matter models involving an s-channel mediator with the ATLAS detector in pp collisions at $\sqrt{s}$ = 13 TeV*. The European Physical Journal C 84 (2024), p.1102. DOI : 10.1140/epjc/s10052-024-13215-5. arXiv:2404.15930

[87] The CMS Collaboration. *Search for new particles decaying to a jet and an emerging jet*. Journal of High Energy Physics 02 (2019), p.179. DOI : 10.1007/jhep02(2019)179. arXiv:1810.10069

[88] The CMS Collaboration. *Search for dark QCD with emerging jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV*. Journal of High Energy Physics 07 (2024), p.142. DOI : 10.1007/jhep07(2024)142. arXiv:2403.01556

[89] Conte, Eric and Fuks, Benjamin and Serret, Guillaume. *MadAnalysis 5, a user-friendly framework for collider phenomenology*. Computer Physics Communications 184 (2013), p.222–256. DOI : 10.1016/j.cpc.2012.09.009. arXiv:1206.1599

[90] Conte, Eric and Dumont, Béranger and Fuks, Benjamin and Wymant, Chris. *Designing and recasting LHC analyses with MadAnalysis 5*. The European Physical Journal C74 (2014), p.3103. DOI : 10.1140/epjc/s10052-014-3103-0. arXiv:1405.3982

[91] Conte, Eric and Fuks, Benjamin. *Confronting new physics theories to LHC data with MadAnalysis5*. International Journal of Modern Physics A33 (2018), p.1830027. DOI : 10.1142/s0217751x18300272. arXiv:1808.00480

[92] Dumont, B. et al. *Toward a public analysis database for LHC new physics searches using MadAnalysis5*. The European Physical Journal C75 (2015), p.56. DOI : 10.1140/epjc/s10052-014-3242-3. arXiv:1407.3278

[93] URL : `http://madanalysis.irmp.ucl.ac.be/wiki/PublicAnalysisDatabase`

[94] de Favereau, J. et al. *Delphes 3: a modular framework for fast simulation of a generic collider experiment*. Journal of High Energy Physics 02 (2014), p.057. DOI : 10.1007/jhep02(2014)057. arXiv:1307.6346

[95] Araz, Jack Y. et al. *Simplified fast detector simulation in MadAnalysis 5*. The European Physical Journal C81 (2021), p.329. DOI : 10.1140/epjc/s10052-021-09052-5. arXiv:2006.09387

[96] The ATLAS Collaboration. *Search for non-resonant production of semi-visible jets using Run 2 data in ATLAS*. Physics Letters B 848 (2024), p.138324. DOI : 10.1016/j.physletb.2023.138324. arXiv:2305.18037

[97] The ATLAS Collaboration. *Search for resonant production of dark quarks in the dijet final state with the ATLAS detector*. Journal of High Energy Physics 02 (2024), p.128. DOI : 10.1007/jhep02(2024)128. arXiv:2311.03944

[98] The CMS Collaboration. *Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s}$ = 13 TeV*. Journal of High Energy Physics 11 (2021), p.153. DOI : 10.1007/jhep11(2021)153. arXiv:2107.13021

[99] The ATLAS Collaboration. *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*. The European Physical Journal C79 (2019), p.970. DOI : 10.1140/epjc/s10052-019-7450-8. arXiv:1907.05120

[100] U. Baur et al. *Excited Quark Production at Hadron Colliders*. Int. J. Mod. Phys. A2 (1987), p.1285. DOI : 10.1142/S0217751X87000661.

[101] G. Altarelli et al. *Searching for New Heavy Vector Bosons in pp Colliders.* Z. Phys. C45 (1989), p.109. DOI : 10.1007/BF01556677.

[102] P. Langacker. *The Physics of Heavy Z' Gauge Bosons.* Reviews of Modern Physics 81 (2009), p.1199–1228. DOI : 10.1103/revmodphys.81.1199. arXiv:0801.1345

[103] Alguero, G. et al. *Signal region combination with full and simplified likelihoods in MadAnalysis 5.* SciPost Physics 14 (2023), p.009. DOI : 10.21468/scipostphys.14.1.009. arXiv:2206.14870

[104] The CMS Collaboration. *Simplified likelihood for the re-interpretation of public CMS results.* (2017). URL : `https://cds.cern.ch/record/2242860?ln=fr`

[105] The ATLAS Collaboration. *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment.* (2019). URL : `https://cds.cern.ch/record/2688062?ln=fr`

[106] Buschmann, M. et al. *Lepton jets from radiating dark matter.* Journal of High Energy Physics 07 (2015), p.045. DOI : 10.1007/jhep07(2015)045. arXiv:1505.07459

[107] Araz, J. et al. *Searches for new physics with boosted top quarks in the MadAnalysis 5 and Rivet frameworks.* The European Physical Journal C 83 (2023), p.664. DOI : 10.1140/epjc/s10052-023-11779-2. arXiv:2303.03427