

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 dans le cadre des modèles de QCD sombre avec le détecteur ATLAS pendant le Run-3 du LHC

Search for emerging jets within dark QCD models with the ATLAS detector during the LHC's Run-3

Présentée par :

Guillaume ALBOUY

Direction de thèse :

Marie-Hélène GENEST

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES

Pierre-Antoine DELSART

CHARGE DE RECHERCHE, CNRS Délégation Alpes

Directrice de thèse

Co-encadrant de thèse

Rapporteurs :

REINA COROMOTO CAMACHO TORO

CHARGE DE RECHERCHE HDR, CNRS DELEGATION PARIS CENTRE

CAROLINE COLLARD

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALSACE

Thèse soutenue publiquement le **15 octobre 2024**, devant le jury composé de :

DAVID ROUSSEAU,

DIRECTEUR DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE SUD

Président

REINA COROMOTO CAMACHO TORO,

CHARGE DE RECHERCHE HDR, CNRS DELEGATION PARIS CENTRE

Rapporteuse

CAROLINE COLLARD,

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALSACE

Rapporteuse

MARINE KUNA,

MAITRESSE DE CONFERENCES HDR, UNIVERSITE GRENOBLE ALPES

Examinatrice

JULIEN DONINI,

PROFESSEUR DES UNIVERSITES, UNIVERSITE CLERMONT AUVERGNE

Examineur

CONTENTS

1	Introduction	6
2	The Standard Model of particle physics	8
2.1	Standard model of particle physics	9
2.2	Quantum Chromodynamics, QCD	11
2.3	Physics in hadron collider	13
3	The dark matter paradigm, dark QCD and Emerging Jets	15
3.1	The dark matter paradigm	16
3.1.1	Evidences of missing matter density	16
3.1.2	Dark matter candidates	17
3.1.3	Detection methods	18
3.2	Composite dark matter in strongly interacting dark sector	21
3.2.1	Motivation for a composite dark matter within a dark sector	21
3.2.2	The specific case of the strongly interacting dark sector	21
3.3	Emerging Jets	24
4	The ATLAS experiment at the CERN Large Hadron Collider during Run-3	27
4.1	The LHC at CERN	28
4.2	A complete description of the ATLAS detector	30
4.2.1	Overview	30
4.2.2	Coordinates system of ATLAS	31
4.2.3	Magnet system	32
4.2.4	Inner tracker	32
4.2.5	Calorimeter system	35
4.2.6	Muon spectrometer	37
4.3	Data taking with ATLAS	38
4.3.1	LHC conditions: luminosity and pileup	38
4.3.2	Trigger system	41
4.3.3	Data taking workflow	42
4.4	Monte Carlo simulation	43
5	Reconstruction of collision events and physical objects with the ATLAS detector	44
5.1	Reconstruction workflow, from raw detector measurements to complete physical reconstruction	45
5.2	Tracks: charged particles trajectory reconstruction	47
5.3	Particle decay vertex reconstruction	51
5.3.1	Primary vertex reconstruction	51
5.3.2	Secondary vertex reconstruction	52
5.4	Jet reconstruction	55
5.4.1	Jet clustering	55
5.4.2	Jet angular opening: small- R and large- R jets	56

5.4.3	Jet input constituents	56
5.4.3.1	Pure calorimeter constituents: topological cell clusters	57
5.4.3.2	Track-calorimeter combined constituents	59
5.4.3.3	True constituents	66
5.4.4	Pileup mitigation and grooming techniques	67
5.4.4.1	Charged Hadron Subtraction	67
5.4.4.2	Neutral constituent subtraction	67
5.4.4.3	Jet grooming algorithms	68
5.4.5	Jet calibration	69
5.4.5.1	Simulation-based calibration	69
5.4.5.2	Data-based calibration: In-situ corrections	72
6	Simultaneous calibration of the energy and mass of large-radius jets using a deep neural network	74
6.1	The standard jet calibration in ATLAS	76
6.2	Samples, reconstruction and selections	79
6.2.1	Event simulation	79
6.2.2	Jet reconstruction and selection	80
6.3	Introduction to Deep Neural Networks	81
6.4	Deep Neural Network calibration methodology	83
6.4.1	Loss function	84
6.4.2	Input features	85
6.4.3	Architecture	87
6.4.4	Training procedure	89
6.4.5	Implementation details	90
6.5	Performances and validation of the calibration	91
6.5.1	Technical performances	91
6.5.1.1	Overfitting	91
6.5.1.2	Convergence	91
6.5.2	Calibration performances	92
6.5.2.1	Response closure and resolution	92
6.5.2.2	Effect of η annotation	94
6.5.2.3	Performances on QCD dijets	94
6.5.2.4	Performance on other topologies	96
6.5.3	Validation of the DNN calibration	99
6.6	Limitations and possible solutions	106
6.6.1	Training convergence and ending criterion	106
6.6.2	Calibration of heavy boosted jets	106
6.6.3	Importance of the features	109
6.7	Conclusion	110
7	Search for emerging jets within dark QCD models with the ATLAS detector during Run-3	111
7.1	Data and Monte Carlo simulations	113
7.1.1	Data	113
7.1.2	Monte Carlo simulations: signal and background	113
7.1.2.1	Background samples	113
7.1.2.2	Signal samples	114

	7.1.2.3	Pileup overlay	116
	7.1.2.4	Monte Carlo re-weighting	116
7.2		Event reconstruction, objects selections and efficiencies	118
	7.2.1	Jets	118
	7.2.2	Tracks and secondary vertices	122
	7.2.3	Special derivation: LLP1	125
7.3		Trigger studies	126
	7.3.1	Muon trigger	126
	7.3.2	Standard large- R jet trigger	127
	7.3.3	Run 3 Emerging Jet trigger	128
	7.3.4	Trigger efficiencies in MC signal samples	130
7.4		Analysis strategy	131
	7.4.1	ABCD data-driven background estimation method	131
7.5		Discriminating observables and data-MC agreement	135
	7.5.1	Signal characteristics	135
	7.5.2	Displaced objects	139
	7.5.3	Jet substructure	145
7.6		Event selections: ABCD plane and signal region definitions	150
7.7		Background estimation, validation of the ABCD method	155
	7.7.1	Nominal test of the ABCD background estimation method in MC	155
	7.7.2	Validation of the ABCD background estimation method in MC	156
	7.7.3	Validation of the ABCD background estimation method in blinded data	159
	7.7.3.1	Data/MC agreement in the ABCD validation region	160
	7.7.3.2	Validation of the ABCD method in data	163
7.8		Systematic uncertainties	167
	7.8.1	Tracking systematics	167
	7.8.2	Jet systematics	168
	7.8.3	Generator systematics	169
	7.8.4	Pileup re-weighting systematic	171
	7.8.5	Total systematic uncertainties in the signal yields in the SR and CRs	171
7.9		Statistical analysis and results	174
	7.9.1	Likelihood-based tests of new physics	174
	7.9.1.1	Test statistics	176
	7.9.1.2	Approximation of distribution of the profile likelihood ratio	179
	7.9.2	Statistical framework	180
	7.9.3	Results and limits on the signal cross-section from MC simulations	181
	7.9.3.1	Nominal test of the statistical framework in MC simulations	181
	7.9.3.2	Predicted upper limits on the signal cross-section	183
	7.9.3.3	Effect of the systematic uncertainties on the signal cross-section upper limits	186
	7.9.3.4	Results in the context of the global search for a new vector mediator Z'	187
7.10		Conclusion	190

8 General conclusion

191

CHAPTER 1

INTRODUCTION

The Standard Model (SM) of particle physics is a very successful theory describing the particle contents and interactions governing the dynamic of the infinitely small in the universe. It is a theory that allowed the prediction of many fundamental particles and the understanding of the fundamental interactions between them, which culminated with the recent discovery of the Higgs boson by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) in 2012. The SM is however acknowledged not to be a complete theory as several inconsistencies and limitations are still present. In particular, the lack of a candidate particle for dark matter is commonly seen as one of its limitations within the particle physics community. Many particle physics experiments thus aim at discovering new particles that could explain the seemingly missing mass density highlighted in the universe. As searches for a new weakly interacting massive particle have not yet produced a discovery, the search for dark matter has recently broadened to more diverse dark matter particles, such as composite particles. Such composite particles are assumed to lay within a distinct and complex dark sector dynamic in analogy to the SM sector. These kinds of theories are very interesting as they are well motivated and yield very exotic signatures in particle physics experiments. In this context, this thesis aims at studying a particular theory of dark sector with a strong dynamic resembling the SM strong interaction, and containing some unstable dark particles with long lifetimes. Such a dark sector is supposed to be accessible through a new massive mediator that can be produced in SM particle collisions and produce a jet-like signature with many displaced objects. This search and this thesis have been conducted with the ATLAS detector at the LHC and with the first proton-proton collision data of Run 3 recorded during the years 2022 and 2023.

First, in Chapter 2, a general introduction to the SM is given and in particular to the strong interaction sector governing the hadronic dynamic. Hadronic colliders are also briefly discussed to provide general knowledge of their operations. Chapter 3 introduces the dark matter paradigm in particle physics with the historical discoveries leading to the concept of a missing matter density in the universe and the following broad query it brought about. This chapter then describes the theory of a composite dark matter within a dark sector and in particular the case of the strongly interacting dark sector. At last, it describes in more details the model of emerging jets explored in this thesis.

Chapter 4 introduces the LHC and the ATLAS detector. A detailed description of the ATLAS detector operation is provided. It is followed by a description of the collision event reconstruction workflow used in ATLAS in Chapter 5. The reconstruction methods of several physics objects of interest for this thesis are discussed.

In Chapter 6, the first project of my thesis is presented which involves the development of a new calibration technique for a certain type of jets using a machine learning method. This new technique is described and motivated, and the obtained performances are compared to the standard calibration technique currently used in the ATLAS Collaboration.

Finally, Chapter 7 presents the second project of my thesis: the search for emerging jets within a strongly interacting dark sector with the ATLAS detector. In this search the dark sector is supposed to be accessible through a new Z' vector boson mediator produced in the s-channel, leading to the production of two exotic jets yielding many displaced objects. The full analysis strategy and workflow is presented and the expected results are shown.

CHAPTER 2

THE STANDARD MODEL OF PARTICLE PHYSICS

Outline of This Chapter

2.1	Standard model of particle physics	9
2.2	Quantum Chromodynamics, QCD	11
2.3	Physics in hadron collider	13

This first chapter aims at introducing the physical context of this thesis. It gives first a succinct introduction to the Standard Model of particle physics (Section 2.1), followed by a description of the Quantum Chromodynamics theory of the strong force (Section 2.2). Some important physical concepts related to hadron colliders are also given in Section 2.3.

2.1 Standard model of particle physics

Progresses in modern particle physics since the second half of the 20th century have been mainly driven by the development of the standard model (SM) [1] jointly with performant physics experiments (such as LEP, Tevatron, HERA, LHC) [2–5] providing precise measurements of particle physics interactions. The SM is a theory that describes the fundamental particles constituting matter and their interactions. It describes three out of four fundamental interactions, the strong, weak and electro-magnetic interactions. It is a computable theory that gives precise predictions on a large range of microscopic matter interactions. These predictions were and are continuously thoroughly tested by experiments. The possibility to test this theory and how well it describes what is observed in experiments is what made the SM the very successful theory of particle physics it is today.

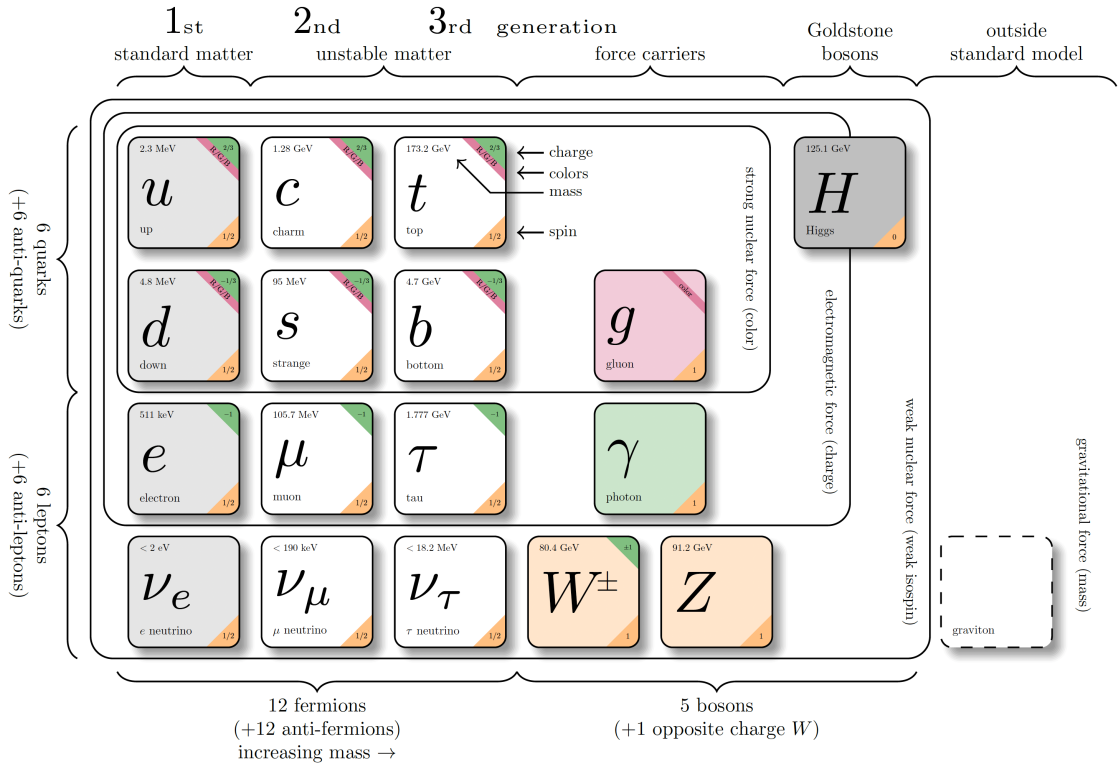


Figure 2.1: Graphic representation of the SM showing the three fermion generations and the bosons corresponding to the three interactions forces along the Higgs boson [6].

The SM is a Quantum Field Theory generalising the Quantum Electro-Dynamics (QED) where particles are interpreted as excitation of quantum fields. It is a renormalisable theory based on local symmetries and a gauge invariance leading to conserved currents and charges. The SM gauge invariance $\mathcal{G}_{SM}$ is a combination of gauge invariances from which arise the strong and electro-weak interactions:

$$\mathcal{G}_{SM} = SU(3) \otimes SU(2) \otimes U(1) \quad (2.1)$$

$SU(3)$ is the gauge theory for the strong interaction called Quantum Chromodynamics (QCD) associated to a colour charge and eight gauge bosons called gluons g interpreted as the strong force carriers. $SU(2) \otimes U(1)$ is the gauge theory for the electro-weak interaction, also called Quantum Flavour Dynamics, associated to

two electro-weak charges (the weak isospin and weak hypercharge) and four gauge bosons W_1 , W_2 , W_3 and B . In the SM, the electro-weak symmetry is spontaneously broken through the Higgs mechanism [7], separating the electro-magnetic and weak interactions. The broken symmetry results in the electro-magnetic $U(1)$ gauge symmetry carried by the massless photon γ and in the actual massive physical particles carrying the weak interaction, $W^\pm$ and Z bosons.

The particle content of the SM is divided between the gauge bosons of spin 1, carriers of the fundamental interactions specified above, and the fermions of spin 1/2, the fundamental particles composing ordinary matter. The SM is completed by the Higgs boson, a scalar boson of spin 0 at the origin of the electro-weak symmetry breaking and providing a rest mass to all fundamental particles (except for gluons, photon and neutrinos which remain massless). The fermions, divided into two families, are characterised by their quantum numbers (charges) and masses. The hierarchy in their masses defines three generations. The first family are the quarks charged under the strong and electro-weak forces and as such interacting with gluons, $W^\pm$, Z and photons. Each generation contains a doublet of quarks which differ by their electric charges. The first generation of quarks are the up and down (u , d), the up carries an electric charge $Q = 2/3$ and the down $Q = -1/3$. Both carry one out of three colour charge making the quarks the only fermions interacting with gluons. The second and third generations of quarks, charm and strange (c , s), and top and bottom (t , b), have the same quantum numbers as the first generation but differs in mass which increases with the generations. The second family of fermions are the leptons, charged under the electro-weak force and as such interacting with $W^\pm$, Z and photons but not with gluons. Similarly to the quarks, the three generations of leptons contain a doublet of leptons with $Q = 0$ or $Q = -1$ electric charge. The electrically charged leptons are the electron e , the muon μ and the tau τ , and to each corresponds a neutral massless lepton called neutrino ν . The three generations of leptons differ by the mass of their electrically charged component. Both for quarks and leptons, only the first generation is stable and compose the ordinary stable matter. The up and down quarks combine to form protons and neutrons and jointly with electrons, they compose the entirety of the baryonic matter (nuclei and atoms) in the universe. Finally, for each fundamental fermion there exists an antiparticle with opposite quantum numbers but same mass. Figure 2.1 gives a graphic representation of the SM with the important properties of the different fundamental particles (electric charges, strong charges, masses and spins).

All these particles were observed and their properties measured by several physics experiments throughout the last century. The last SM particle to be discovered was the Higgs boson in 2012 by the ATLAS and CMS experiments [8, 9]. However, while being very successful experimentally and theoretically, the SM has known flaws and limitations, among others: the neutrinos' experimental masses, the strong CP problem, the fourth fundamental interaction of gravity is not included, the origin of dark matter remains unexplained. It is thus acknowledged that it is not a complete theory but rather an excellent low energy approximation to a more fundamental theory. The dark matter problem will be discussed further in Section 3.1.

2.2 Quantum Chromodynamics, QCD

The strong interaction is described by the Quantum Chromodynamics field theory (QCD) in the SM [10]. It describes the interactions between the colour-charged quarks, carried by the gluons, and the interactions amongst the gluons themselves. In the SM, the strong sector counts three colours (Red, Green and Blue) and eight gluons, such that each quark has three identical colour replicas and interact together through the exchange of gluons represented as colour exchanges, each gluon carrying a colour-anticolour charge. The QCD results from the $SU(3)$ gauge symmetry, a non-abelian symmetry, where the quark fields are elements of the fundamental representation of $SU(3)$ and the gluon fields are elements of the adjoint representation of $SU(3)$. From this symmetry arises specific properties differing from the electro-magnetic and weak interactions.

First, gluons are colour-charged particles such that they are self-coupled. Therefore, the physical interactions (vertices) of QCD are not restricted (as in QED) to the boson-fermion-antifermion interaction with the gluon-quark-antiquark vertex analogous to the QED photon-electron-positron vertex but also include the gluons self-interactions with 3-gluon and 4-gluon vertices. The vertices of QCD are represented in Fig. 2.2. This property allows both quarks and gluons to radiate gluons when propagating.

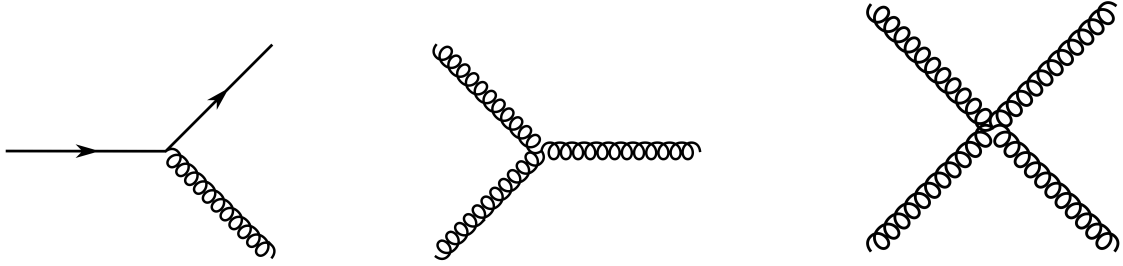


Figure 2.2: Interaction vertices of QCD. Black lines represent quarks and strings represent gluons [11].

Second, as for QED, the coupling constant (the force strength) of QCD, α_S , is not actually a constant as it varies with the energy scale of a given interaction Q^2 . For a given interaction, Q^2 is taken as the squared four-vector transferred during the interaction. $\alpha_S(Q^2)$ is generally expressed as:

$$\alpha_S(Q^2) = \frac{1}{\beta \ln(\frac{Q^2}{\Lambda_{\text{QCD}}^2})}$$

where β is a known constant and Λ_{QCD} is an energy scale of order of few hundred MeV. Contrary to QED where $\alpha_{\text{QED}}(Q^2)$ increases with Q^2 (i.e. increases at short distance), the QCD running coupling $\alpha_S(Q^2)$ decreases with increasing Q^2 , meaning that for hard, energetic processes, or equivalently at short distances, the strength of the strong force is small. The distribution of α_S as a function of Q is shown in Fig. 2.3 as a result of numerous measurements.

This property of QCD implies two interesting phenomena, the asymptotic freedom and confinement of quarks. In hard-scatter processes, the QCD becomes weak allowing the free propagation of quarks and gluons regarding the strong force (asymptotic freedom). In contrast, at lower energy, with $Q^2 \leq \Lambda_{\text{QCD}}^2$,

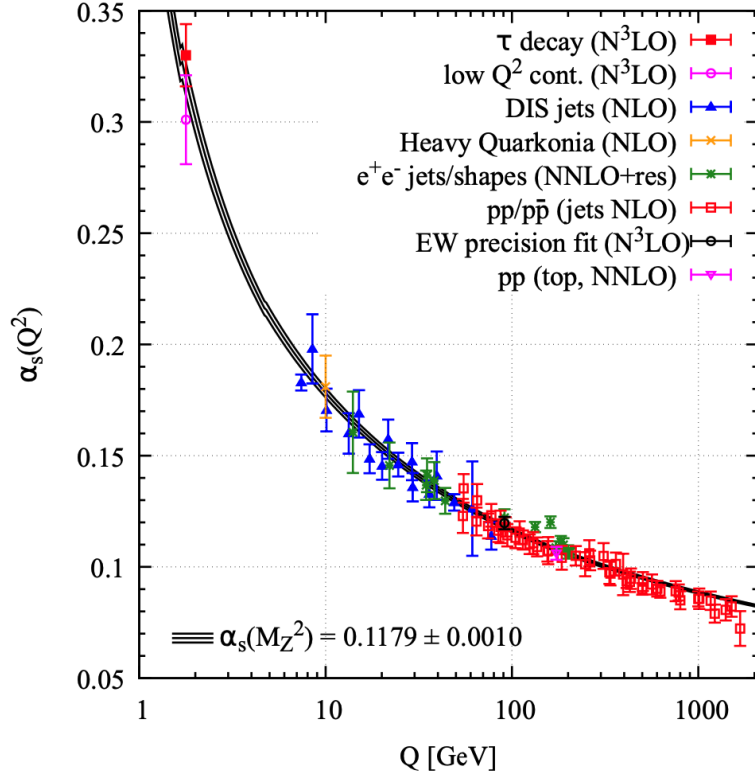


Figure 2.3: Summary of measurements of $\alpha_s(Q^2)$ from multiple type of interactions and using different degrees of QCD perturbation theory in the extraction of α_s (NLO: next-to-leading order; NNLO: next-to-NLO; NNLO+res: NNLO matched to a resummed calculation; N³LO: next-to-NNLO) [12].

the strong force's strength becomes very large, leading to confinement of quarks and gluons in tightly bound composite colour neutral states called *hadrons*. These hadrons are colour singlets and form two families of particles, the baryons (proton, neutron, ...) composed of three quarks and mesons (pion, rho, ...) composed of a quark-antiquark pair. Moreover, due to this property of confinement it is not possible to separate colour charges and to observe individual quarks or gluons when they are emitted in high-energy collisions. An emitted high-energy quark or gluon with energy Q will be first in a state of asymptotic freedom where $Q^2 \gg \Lambda_{QCD}$, propagating and losing energy by radiating gluons or pair-producing quarks during what is called a *parton shower*. The parton shower then stops when the initial quark or gluon energy becomes comparable to Λ_{QCD} , $Q^2 \leq \Lambda_{QCD}$. The produced quarks and gluons subsequently confine in pairs of quark-antiquark or triplet of quarks, forming colourless hadrons. This stage is called *hadronisation*. The obtained object from the emission of an energetic quark or gluon is a flow of collimated hadrons called a *jet*.

The QCD coupling α_s is also, contrary to QED or weak interaction, large in an extended range of energy such that perturbative approach generally cannot be used for QCD theoretical calculations in this regime. For example, at an energy corresponding to the mass of the Z boson, $\alpha_s = 0.1183 \pm 0.0009$ [13]. In this regime, a non-perturbative approach can be used called lattice QCD [14]. However, in physics colliders where the interaction energies are large, perturbative theory can in practice be used for hard-scattering process calculations but corrections to the leading order of perturbative theory (NLO: next-to-leading order; NNLO: next-to-NLO; ...) are required for precise predictions.

2.3 Physics in hadron collider

Modern hadron colliders mainly use highly-energetic (TeV scale) proton-proton (LHC) or proton-antiproton (Tevatron) collisions to probe the SM. At these energy scales, hadron collisions probe lower distance scales than the hadron's size, such that it is not the hadrons themselves that interact but the fundamental particles (quarks and gluons), usually called *partons*, composing them. These partons carry a fraction of the hadron energy and multiple parton interactions are possible during a single hadron-hadron collision. Generally, interactions between two energetic (hard) partons are the most interesting for physics analyses as they allow to probe hard-scattering (HS) processes with the production of energetic particles (e.g. quark or gluon with a large transferred energy, or boosted massive particles like $W^\pm$) with large deflection angles. It is to be opposed to low energy (soft) interactions, elastic scattering with low transferred energies and low deflection angles with the collision line. However, due to the composite nature of hadrons, multiple phenomena can happen during a single collision beside the HS interaction. First, as mentioned, multiple parton interactions are possible. In addition to the HS process, secondary softer HS parton interactions originating from the same incoming hadrons are likely to occur. These secondary interactions lead to lower energy production of particles but with important deflection angle such that these interactions contaminate the HS final states. These secondary interactions are usually referred as the *underlying event* (UE). Second, the partons from the incoming hadrons can emit partons before the collision, called initial state radiation (ISR). Partonic emissions can also be produced after the collision by the final state particles of the HS process, called final state radiation (FSR). Both ISR and FSR are usually sources of contamination for the HS event but they can also be a handle in exotic analyses (e.g. in invisible dark matter searches). This is represented in Fig. 2.4 showing a graphic representation of a hadron collision involving a HS interaction producing partons and the underlying event.

Hadron colliders are excellent experiments for QCD physics as they are probes of the hard-scattering of partons such that a very large proportion of HS events and soft-scattering lead to the production of jets. For this reason, jets are very frequently used in physics analyses or they can compose the main background events. However, the physics analyses also suffer from the intricate QCD dynamic of parton interactions that increases the complexity of the final states.

Finally, incoming hadrons travel on the same axis, such that the hadrons do not have momenta transverse to this axis. This imposes, because of momentum conservation, that the final state particles have a total transverse momentum value of zero (i.e. a transverse momentum balance). This constraint on the momentum balance can then be harnessed to detect missing energy in the final states, generally referred as missing energy transverse (MET), that can come from particles such as neutrinos, which interact only weakly and are not detected.

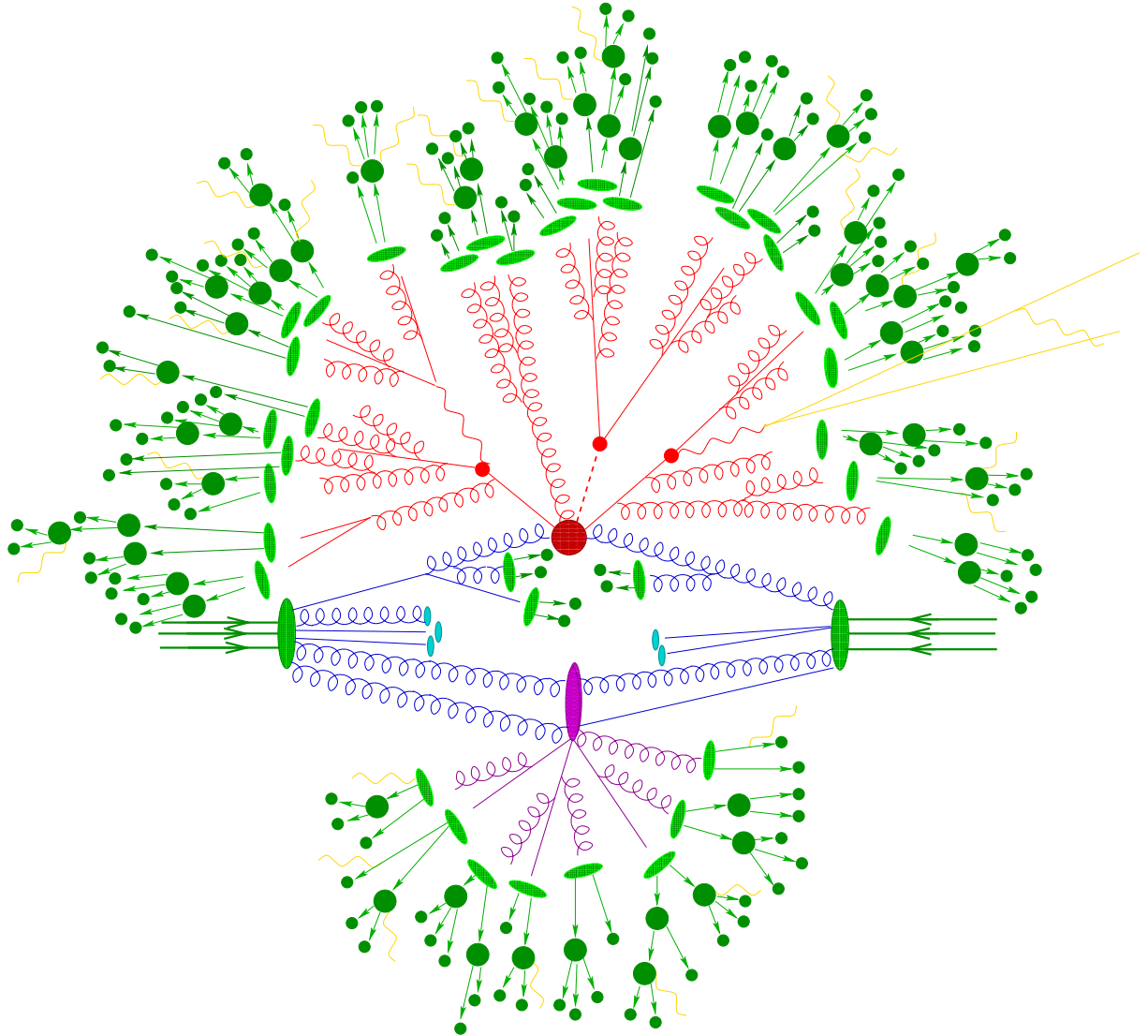


Figure 2.4: Schematic of a hadron-hadron collision leading to parton emissions. The red blob is the HS interaction surrounded by the resulting parton shower. The purple blob represents the UE from the same hadron-hadron collision. The red lines represent the parton shower. Hadronisation is represented by the light green blobs while the dark green indicate hadron decays. The yellow lines are soft photon radiations [15].

CHAPTER 3

THE DARK MATTER PARADIGM, DARK QCD AND EMERGING JETS

Outline of This Chapter

3.1	The dark matter paradigm	16
3.2	Composite dark matter in strongly interacting dark sector	21
3.3	Emerging Jets	24

In this second chapter, the dark matter paradigm is introduced as the main motivation for this thesis work. A particular model of strongly interacting dark sector is presented and is the focus of a physics analysis within the ATLAS experiment for which I am one of the main analysers. The first Section 3.1 presents the historic foundations of the dark matter problem in the 20th century and the paradigm it has become in modern physics and in particular in particle physics. Then Section 3.2 discusses the theoretical aspects of a composite dark matter within a strongly interacting dark sector. Finally, Section 3.3 outlines the theoretical model of Emerging Jets which is at the centre of this thesis physics analysis.

3.1 The dark matter paradigm

Dark matter (*DM*) is a hypothetical matter field that would make up for 85% of the total mass content of the Universe. The existence of this exotic matter was first theorized to account for the observed missing matter density in galaxies and large structures of the Universe [16]. From unknown origin, DM is one of the great mystery of modern physics in the diverse domains of cosmology, astrophysics and particle physics. As of today no experimental observations explaining the nature of DM have been obtained.

3.1.1 Evidences of missing matter density

The first evidence of missing matter density in the Universe was observed by Zwicky in 1933 from the measurement of velocity dispersion in rich galaxy clusters [17]. In order to keep bound states in such clusters he estimated that ten to a hundred times more mass than what could be observed through galaxy luminosities was required. He thus proposed the existence of an invisible matter density (i.e. DM) that would account for these discrepancies. Several papers followed his work either questioning the calculations or showing similar results, such that the DM hypothesis was not commonly accepted nor rejected at this time, showing the need for more observations.

The second largely acknowledged evidence for DM came from the measurements of galactic rotation curves by several astrophysicists in the late '70s which definitely hardened the DM hypothesis. The main historical references are the works of V. Rubin and K. Freeman [18, 19] with the observations of M31 and M33 galaxies. Using Newtonian gravitational dynamics, it is possible to show that for stars in spiral galaxies their circular velocity's radial distribution (commonly referred as rotation curve) is directly related to the mass distribution of the galaxy. By only accounting the mass from observable luminous stars, the expected rotation curves should scale according to $1/\sqrt{r}$ (with r the radial distance from the centre of the galaxy). However, measurements from Rubin and Freeman showed unexpected flat rotation curves at large distance from the galactic centre. The results on the observed galactic rotation curves implies that the mass distributions of spiral galaxies peaks at larger radii than what is expected from the luminous signals. This presented the need for an additional invisible matter density as large as the mass of the detected galaxies with a very different mass distribution. These results were followed by many similar observations in other galaxies. The common hypothesis for this additional matter is in the form of halos of DM surrounding the galaxies. An example of observed rotation curve is shown in Fig. 3.1, along with the contributions from the visible (stars) components, the gas components (observable through X-rays) and the expected contribution from a halo of dark matter around the galactic centre. From these observations, the need for an invisible DM which does not interact electro-magnetically but has a gravitational interaction became increasingly acknowledged by the astrophysics community.

A third astrophysical evidence came from a system of colliding clusters of galaxies, called the Bullet Cluster [21]. This particular system allows to study the behaviour of the different matter components in the clusters. The first component are the stars within the different galaxies which are slightly affected by the collision and thus mostly traverse freely during the collision. Second, the hot gases (observable through X-rays) which compose most of the baryonic mass of the two clusters, are strongly affected during the collision due to their electro-magnetic interactions. This leads to the gases from the pair of clusters to slow down much more and to concentrate at the centre of the collision. Using gravitational lensing¹ it

¹ Principle of observing the bending effect of a massive object on the light emitted by background objects in order to study its mass distribution.

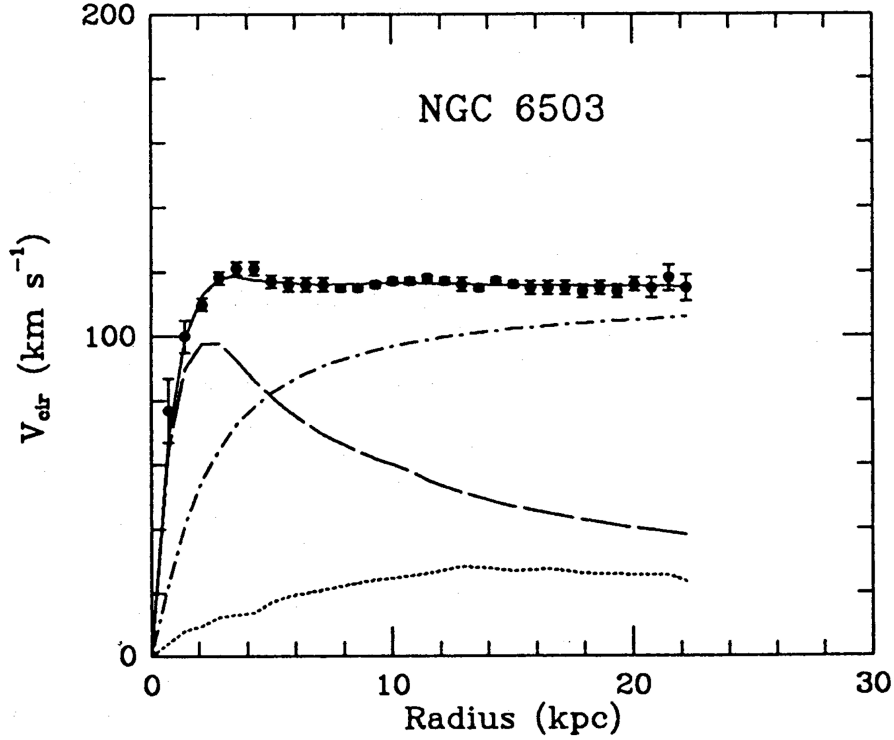


Figure 3.1: Measured rotation curve of the galaxy NGC 6503 (i.e. circular velocity versus the radial distance from the centre of the galaxy) from Ref. [20]. The measurements are represented by the black dots. In addition, the contributions from the visible (dashed curve), gas (dotted curve) components and the expected contribution from a dark matter halo (dash-dot curve) are also shown.

was observed that the strongest lensing effects are not correlated to the hot gas distribution (the dominant baryonic mass component) but follows closely the star components from the two clusters. This can be explained by the presence of two massive DM regions originating from the two clusters and which were unaffected by the collision other than gravitationally, similarly to the stars. Similar observations were made for other colliding clusters systems [22].

DM can also be summoned in cosmology in the Lambda Cold Dark Matter (Λ CDM) model [23] which describes the DM contributions to the Cosmic Microwave Background (CMB) anisotropies and baryonic oscillations and the formation of large scale structures in the Universe in agreement with experimental observations, including the Planck's measurements of the CMB [24]. These diverse evidences for gravitational contributions from an invisible matter density from astrophysics, but also cosmology, are piling up at the door of physicists, making the DM paradigm a major inquiry for modern physics and as of today the nature of DM remains a mystery.

3.1.2 Dark matter candidates

As the term DM historically referred to any "dark" materials including astrophysical materials too faint to be detected, the first most obvious candidates to physicists for DM were dense objects composed of baryonic matter that does not emit enough light to be observed. Multiple types of objects were proposed like planets, dwarf or dead stars, neutron stars or black holes that would form a halo around the galaxies. These were

referred as the MACHOs (for Massive Astrophysical Compact Halo Objects). This hypothesis was rapidly disfavoured as no indirect observations through gravitational micro-lensing² led to an estimation in the number of MACHOs which could sufficiently accounts for all the missing mass. It is today acknowledged that the MACHOs make up for only a very small portion of the DM and is therefore not a candidate to explain the bulk of DM any more.

By the late '90s it had become clear that the DM were not compact objects but instead new unknown massive particle species that can only, if at all, weakly interact with the SM particles and which were produced at the same time as SM particles during the Big Bang. A large portion of the search for DM particles revolved around the WIMP paradigm (for Weakly Interacting Massive Particle). It describes a kind of particle that weakly interacts with the SM (with a strength below or similar to the SM weak interaction) but is massive and thus interacts gravitationally. From astrophysical and cosmological observations, WIMPs must have some standard properties, they need:

- to be stable or at least with a lifetime of the order of the age of the Universe as evidences for DM appears both in the early Universe (in the CMB) and nowadays (through gravitational interaction);
- to have a small enough self-annihilation cross-section so that the quantity of DM remains stable along the Universe evolution;
- to be non-relativistic ("slow moving", cold) for small scale structures (galaxies, small clusters) to form first during the evolution of the Universe, this property leads also WIMPs to be unable to overcome their mutual gravitational effect, creating clumps of WIMPs (e.g. galaxies DM halos instead of scattered DM).

Many theoretical models propose possible WIMPs, predicting their defining properties (mass, self-annihilation cross-section, couplings to SM particles), which then can be searched for by experiments either in astrophysics or particle physics. From the many proposed models, supersymmetry (SUSY) [25] was one of the most promising model proposing WIMPs, but after a decade of search at the LHC, no evidence for SUSY has been observed [26]. Many experiments searched for WIMPs without finding any evidence leading to important constraints on the WIMP properties [27]. As of today the unconstrained phase space of the WIMPs is greatly reduced. This led physicists to imagine more complex candidates for DM such as: axions [28] or composite dark matter. This second example is the focus of this thesis and will be discussed further in the next section.

3.1.3 Detection methods

As the DM searches revolved around the pursuit for new particles, three main experimental methods were developed built on the DM interactions with itself or with SM particles. A graphic representation of these methods is shown in Fig. 3.2.

On the astrophysics side, the indirect detection is harnessed. It revolves around the hypothesis that DM particles can annihilate into SM particles (photons, electrons, neutrons, . . .) which can then be detected during sky observations. Generally, the searches consist in looking for excesses in the cosmic rays fluxes reaching Earth that would be evidences for DM-generated cosmic rays and would be correlated to the DM properties (mass and self-annihilation cross-section).

² Based on a similar principle as the gravitational lensing, micro-lensing looks for shifts in brightness induced by small transient objects like stars or planets.

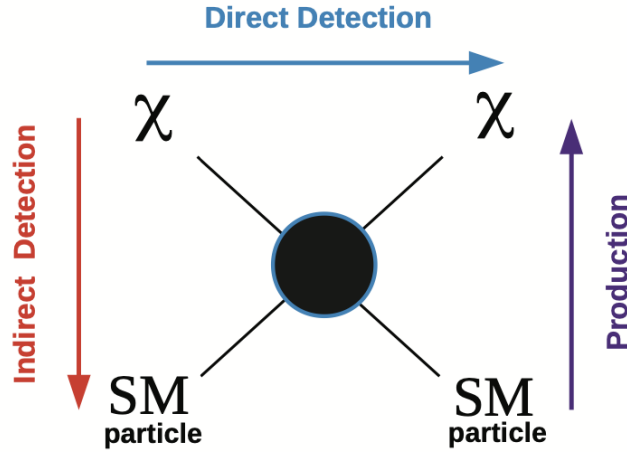


Figure 3.2: Graphic representation of the DM experimental detection methods.

On the particle physics side, there are two methods explored: the direct detection and detection through DM production. The direct detection method is based on the weakly interacting aspects of the DM. It supposes that Earth is continuously crossed by DM particles from the Milky Way DM halo and that DM particles can interact with SM particles in a detecting device. For WIMPs, this means interacting with atomic nuclei or electrons in dense medium through the weak interaction. In that case, the principle is to monitor a very well shielded block of matter for atomic or electronic recoils produced by DM-SM interactions. The shielding is used to reduced as much as possible contamination from SM particles that constitutes the main background for these experiments. The production method supposes that SM collisions can lead to DM particle production. This method was and is actively explored in particle colliders and the associated detectors like the LHC. These collisions would lead a priori to an invisible final state as the produced DM particles would very much likely not interact with the detectors. Collider searches are thus generally based on the measurement of large missing transverse momentum (MET), as a result of the invisible produced DM particles, and associated with one or more SM visible particle probes. This is represented in Fig. 3.3 along a possible DM production mechanism. The accompanying visible particles, which are needed for detection of the collision, can either be produced in the same process as the DM particles or be an initial state radiation (ISR).

All these detection methods aim at measuring or constraining the DM particle properties and are complementary. As of today no experiments has found a significant excess that could be confirmed by similar experiments. However, they allowed to place important constraints on the possible properties of DM particles.

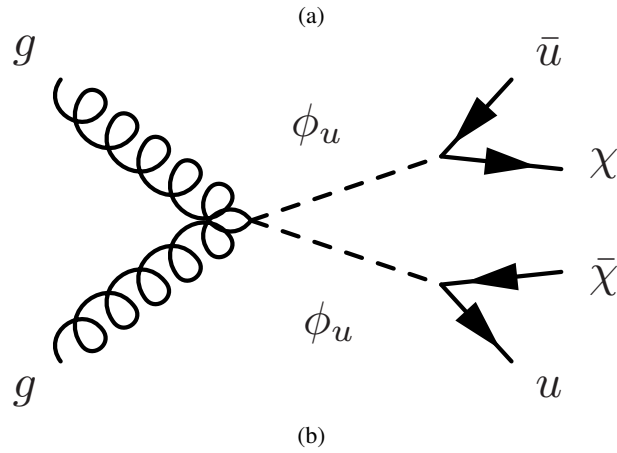
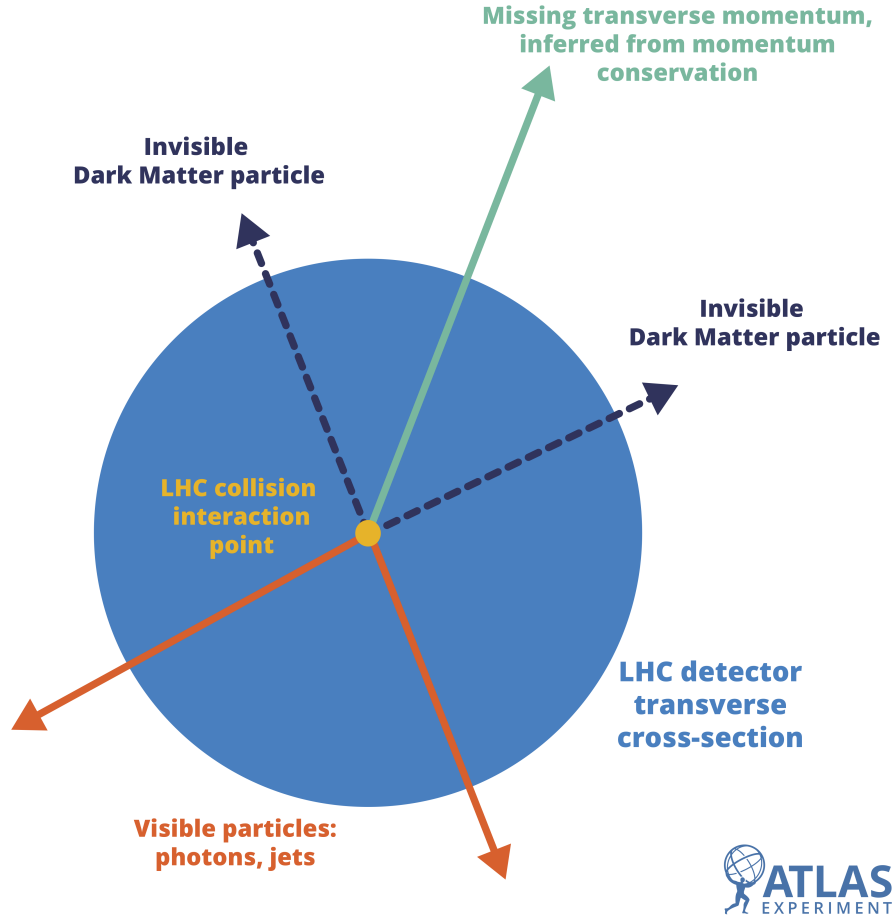


Figure 3.3: (a) Graphic representation of a DM production in association with SM visible particles in a proton-proton collision [29]. (b) Example of a Feynman diagram for the production mechanism of DM particles in association with two jets through a hypothetical new scalar mediator ϕ_u [30].

3.2 Composite dark matter in strongly interacting dark sector

The search for WIMPs as DM candidates has been the focus of most of the astrophysics and particle physics experiments for several decades. However, since no evidences for WIMPs have been observed so far, there has been important theoretical and experimental developments to go beyond the WIMP paradigm and to seek for more complex DM candidates, among which composite DM is a strong suitor.

3.2.1 Motivation for a composite dark matter within a dark sector

Dark sectors are derivative of the asymmetric DM (ADM) paradigm. The ADM model [31] proposes DM candidates as the stable members of a relatively complex additional gauge theory constituting of a SM-decoupled dark sector. The ADM is motivated by the apparent similar mass density (Ω) between the baryonic matter and DM:

$$\Omega_{\text{DM}} \simeq 5 \times \Omega_{\text{baryon}}$$

This similarity, appearing as a coincidence in the WIMP paradigm, suggests a common origin and a strong connection between the production mechanism and the cosmological evolution for the baryonic and dark matter. As the actual mass density of the baryonic matter is the result of the matter-antimatter asymmetry in the early Universe, it is well motivated that the DM density is similarly due to a DM particle-antiparticle asymmetry. The baryonic and DM asymmetries would then be correlated through processes in the early Universe before being decoupled.

Moreover, the baryonic content of the Universe is made of a few number of SM matter fields (protons, bound neutrons and electrons) which are the stable remnants of a larger SM sector. Similarly, DM could be the stable remnants of a larger dark sector resulting from an additional gauge theory $\mathcal{G}_{\text{DM}}$. This introduces sheer complexity to DM models but results in very interesting phenomenology. This gauge theory $\mathcal{G}_{\text{DM}}$ can have similar fields and behaviours as the SM gauge theory (QCD, symmetry breaking, . . .). The resulting dark sector can contain stable and unstable particles, the former ones being excellent DM candidates, in the form of complex composite structures like dark hadrons or dark atoms.

Dark sectors are well motivated and provide a vast phenomenology to explore for physics experiment and in particular collider physics if one supposes that the dark sector is accessible from the SM through a portal.

3.2.2 The specific case of the strongly interacting dark sector

The searches for DM at the LHC have mainly focused on WIMPs, but without any evidences so far the LHC collaborations have started to look for alternative models that would yield exotic signatures that have not been explored yet. Dark sectors are perfect candidates and in particular the models of strongly interacting dark sectors generally referred as dark QCD [32]. The dark QCD models propose that the additional dark gauge group $\mathcal{G}_{\text{DM}}$ is a non-abelian gauge group $SU(N_{\text{CDM}})$ leading to a SM QCD-like dark sector.

Similarly to the SM QCD, this dark QCD extension has a complex structure. It contains matter fields (dark quarks q_D) and gauge fields (dark gluons) but without any prior expectation on the gauge group dimension (the number of dark colours N_{CDM}) or number of matter fields (the number of dark flavours N_f). As for the

SM QCD, dark quarks confine below a confinement scale Λ_D and show asymptotic freedom with a running coupling constant decreasing with the energy. This leads dark quarks and gluons to confine in dark hadrons and, when emitted at high energy, to undergo a parton shower and hadronisation, thus forming dark jets (in analogy with SM jets). These dark hadrons can either be stable, providing DM candidates, or unstable, depending on the symmetries of the dark sector. Finally, this dark sector is supposed to be accessible from the SM through a portal which can either be a boson mediator like the Higgs, a new vector boson Z' , or a new scalar bi-fundamental mediator ϕ charged under both the SM and dark sector symmetries³. These models of dark QCD are very interesting for collider physics as they provide several exotic signatures depending on the model parameters.

At the LHC, the current focus is on models with $\Lambda_D \ll \sqrt{s}$ (with $\sqrt{s}$ the centre-of-mass energy of the collision) and the production of dark quarks either through a Z' (s-channel) or a bi-fundamental ϕ (t-channel). Such produced dark quarks generate dark jets composed of dark hadrons. These dark hadrons are either stable and stay within the dark sector or can decay back to the SM through the same portal. Depending on the fraction of stable dark hadrons (the invisible component) and their decay lifetime, the produced signatures in the particle detectors can be very distinct from SM QCD jets. Fig. 3.4 is a diagram presenting the different topologies of dark jets depending on the dark hadron lifetime and the fraction of stable dark hadrons (invisible) particles in the jets. There are three main dark jets topologies:

- The QCD-like dark jets [33] (bottom left corner in Fig. 3.4) obtained in models where all the dark hadrons decay back to the SM with a negligible lifetime (i.e. prompt decay), leading to fully visible dark jets. This is the topology that is the less distinct from SM QCD jets but still yields distinguishable properties in the jets' internal structure.
- The semi-visible dark jets (SVJ) [34, 35] obtained in models with a non-negligible fraction of dark hadrons that do not decay to the SM (the Exotic (II) region in Fig. 3.4). They are referred as semi-visible jets due to a significant fraction of invisible, dark hadrons, component remaining in the jets. In case of pair produced SVJs, these models produce a signature with MET aligned with one of the jets and particular jet internal structure.
- The emerging dark jets (EJ) [36, 37] obtained in models with fully decaying dark hadrons which have a significant lifetime leading to displaced objects in the jets (the Exotic (I) region in Fig. 3.4). This topology is the most distinct from SM QCD jets. EJs are the main subject of this thesis and such will be fully detailed in the next section.

The dark jets parameters phase space also contains a WIMP-like region where the dark jets are completely invisible either due to very large fractions of stable dark hadrons (bottom right corner in Fig. 3.4) or due to very large dark hadron lifetimes leading to collider stable dark hadrons (top left corner in Fig. 3.4) or both (top right corner in Fig. 3.4). In this region the produced dark jets are undistinguishable from WIMP scenarios and the results from WIMP searches thus apply.

While the dark QCD models provide candidates for DM, the actual main motivation of the LHC searches are in the exotic signatures they produce which are/were unexplored yet and which provide smoking guns for beyond the SM physics (BSM). Therefore the LHC searches are signature-driven without strong theory priors on the model parameters and the first objective is the BSM potential discovery in itself.

³ This bi-fundamental mediator ϕ decays to a pair SM quark-dark quark, $\phi \rightarrow qq_D$.

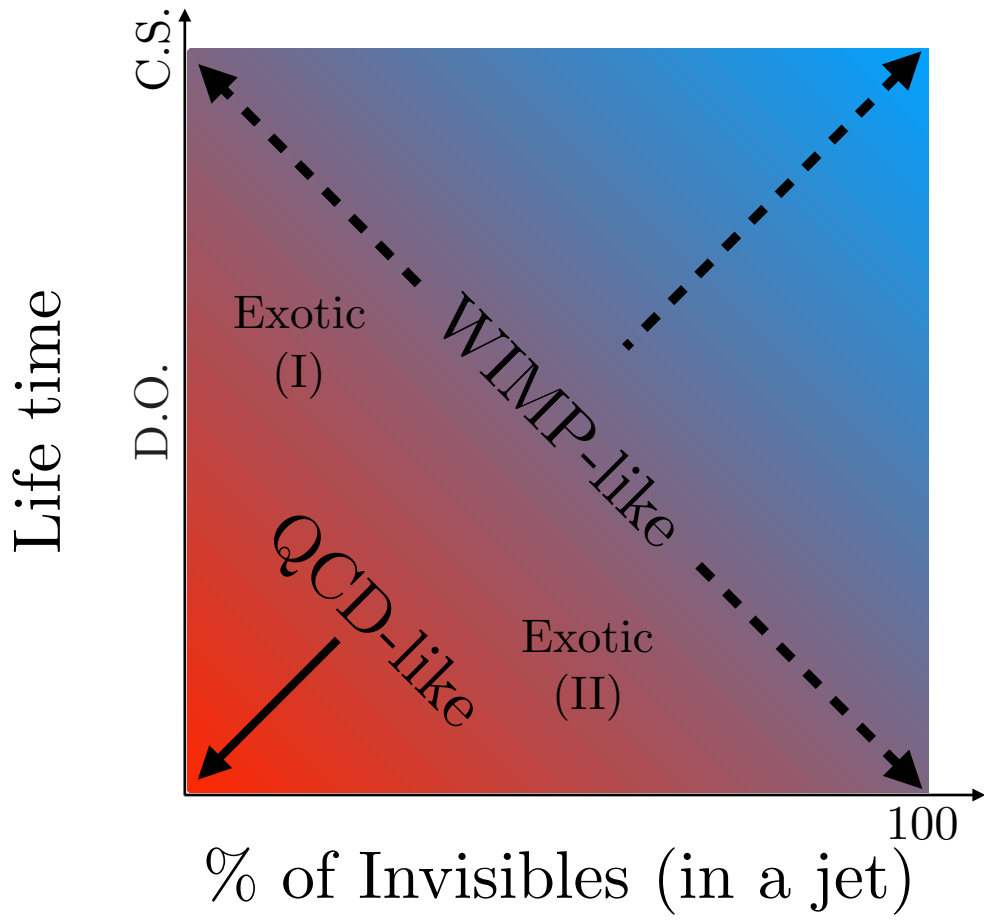


Figure 3.4: Diagram of the different dark jet signatures as a function of the ratio of invisible particles and the decay lifetime of the dark hadrons [38]. C.S. refers to "collider stable" and D.O. to "displaced objects".

3.3 Emerging Jets

Within the dark QCD models, if there is a large mass hierarchy between the portal mediator (Z' or ϕ) and the dark sector mass scales, the decay of the dark hadrons to the SM can be quite slow such that they can travel macroscopic distances before decaying. This is the main motivation for the EJ topology: a particular set of theory parameters can lead dark hadrons to have large lifetimes leading to macroscopic decay lengths [39]. In this scenario, the produced dark jets will gradually turn visible as the dark hadrons progressively decay back to the SM after travelling some distances. Such dark jets seem to emerge into the visible sector. In particle colliders and detectors, EJs are a very exotic signature that is schematically shown in Fig. 3.5: no visible particles are produced at the collision point, detectable (visible) particles only appear further away from the collision and the dark hadron decays produce displaced decay vertices. This aspect is what makes the EJs a very interesting exotic signature as in the SM there are very few particles with macroscopic decay lengths which have the potential to produce displaced vertices⁴. The EJs thus are part of the broader field of BSM searches for long-lived particles (*LLPs*).

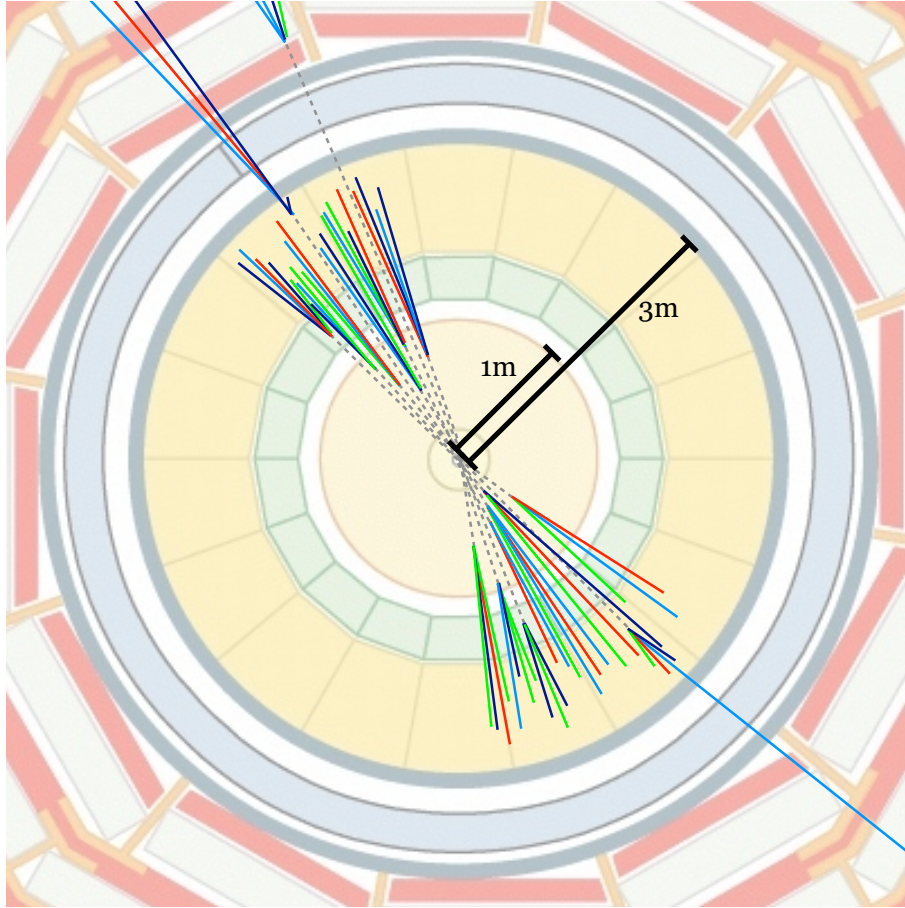


Figure 3.5: Depiction of a pair of EJs produced during a SM collision within a transverse cross-section of a generic cylindrical particle detector composed of a particle tracker at the centre (up to a radius of 1m) and a calorimeter (above 1m). Dashed lines represent the invisible dark hadrons before decaying to the SM and the coloured lines represent their SM decay products [39].

⁴ Mainly B-hadrons and C-hadrons.

The model studied in this thesis is based on the models discussed in Ref. [39]. It supposes a dark QCD sector accessible from the SM through a new vector mediator Z' . The Z' couples to both SM quarks q and dark quarks q_D with the respective coupling constants g_q and g_D . The Z' originates from a $U(1)'$ symmetry broken at the TeV scale such that

$$\mathcal{G}_{\text{DM}} = SU(N_{\text{CDM}}) \times U(1)'$$

The process of interest is the following one: production of a Z' mediator through a quark-antiquark $q\bar{q}$ annihilation, which then decays to a pair of dark quark-dark antiquark $q_D\bar{q}_D$:

$$q\bar{q} \rightarrow Z' \rightarrow q_D\bar{q}_D$$

The corresponding Feynman diagram is given in Fig. 3.6.

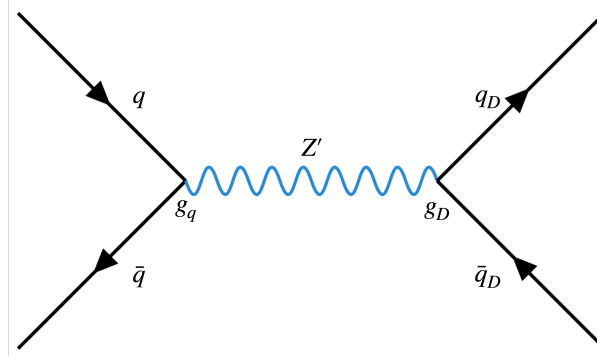


Figure 3.6: Feynman diagram for the pair production of dark quarks.

This process produces two EJs emitted back-to-back composed of many displaced decay vertices as represented in Fig. 3.5.

In the dark QCD sector, the dark baryons carry a conserved dark baryon number such that the lightest one is considered stable and constitute the DM candidate of the theory. On the other hand the dark mesons are not charged under this dark sector symmetry and thus can decay back to the SM. It is supposed that only the lightest dark meson, referred as the dark pion π_D (in analogy with the SM pion), decay to SM in a pair of down quarks:

$$\pi_D \rightarrow d\bar{d}$$

where the produced SM quarks then undergo their own hadronisation in the SM sector. Thus, the produced visible particles forms a jet in the detectors which, in terms of constituting particles, is similar to SM jets (i.e. contain mainly SM hadrons). However, due to the double hadronisation, first of the dark quarks in the dark sector and secondly of the SM quarks, the EJs are generally significantly broader than SM jets.

The dark pion lifetime $c\tau$ is given by

$$c\tau \approx 80\text{mm} \times \frac{1}{g_q^2 g_D^2} \times \left(\frac{2\text{GeV}}{f_{\pi_D}} \right)^2 \left(\frac{100\text{MeV}}{m_{\text{down}}} \right) \left(\frac{2\text{GeV}}{m_{\pi_D}} \right) \left(\frac{m_{Z'}}{1\text{TeV}} \right)^4 \quad (3.1)$$

where f_{π_D} is the dark pion decay constant, m_{down} is the SM down quark mass, m_{π_D} is the dark pion mass and $m_{Z'}$ is the Z' mass. Finally, it is supposed that all the dark hadrons in the EJs are dark mesons and that all the dark mesons decay promptly to dark pions.

The main parameters of the model are:

- $m_{Z'}$, the mass of the vector mediator Z' ,
- g_q , the coupling constant of the Z' to SM quarks,
- g_D , the coupling constant of the Z' to dark quarks,
- N_{CDM}, N_f , the number of colours and flavours of the dark sector gauge group $SU(N_{\text{CDM}})$,
- Λ_D , the confinement scale of the dark sector,
- m_{q_D} , the dark quark mass,
- m_{π_D} , the dark pion mass,
- $c\tau$, the dark pion lifetime.

There are no strong constraints from theory priors except for the required large mass hierarchy between the Z' ($m_{Z'}$) and the dark sector mass scale (Λ_D), the TeV to GeV hierarchy is favoured to have millimetres to metres dark pion decay lengths. Moreover, it is well motivated to have a dark sector at the GeV scale, comparable to the SM, as similar production mechanisms are inferred for DM and baryonic matter in the ADM paradigm.

CHAPTER 4

THE ATLAS EXPERIMENT AT THE CERN LARGE HADRON COLLIDER DURING RUN-3

Outline of This Chapter

4.1	The LHC at CERN	28
4.2	A complete description of the ATLAS detector	30
4.3	Data taking with ATLAS	38
4.4	Monte Carlo simulation	43

The ATLAS detector [40] is one of the four main detectors of the Large Hadron Collider (LHC) with CMS [41], ALICE [42] and LHCb [43]. It is a general purpose detector dedicated to precisely study the SM and search for BSM physics. The ATLAS detector collected large quantities of high-quality data that led to a number of physics results and to the discovery of the Higgs boson in July 2012 [44]. All those results pushed the precision levels of the SM measurements. However, the SM suffers from limitations (as discussed in Chapters 2 and 3) that are still unanswered and some of them are thoroughly investigated by ATLAS.

This chapter first introduces and presents the LHC and the CERN accelerators complex (Section 4.1). Then a complete description of the ATLAS detector is given (Section 4.2), focusing more on the sub-detectors that are of interest for this thesis work. It is followed by the description of the data taking system of ATLAS (Section 4.3). Finally, a short section presents the Monte Carlo simulation process of proton–proton collisions within the ATLAS detector (Section 4.4).

4.1 The LHC at CERN

The LHC is a particle collider situated at CERN at the France-Switzerland frontier, near Geneva. It is the largest and most energetic hadron collider in the world, able to produce proton–proton collisions and heavy-ion collisions (mainly lead–lead). In this chapter and the following ones, only the proton–proton (pp) collisions will be discussed. The LHC is at the end of the CERN’s long accelerator complex allowing to get pp beams with very high collision rate and high energy. Figure 4.1, presents the CERN accelerator complex.

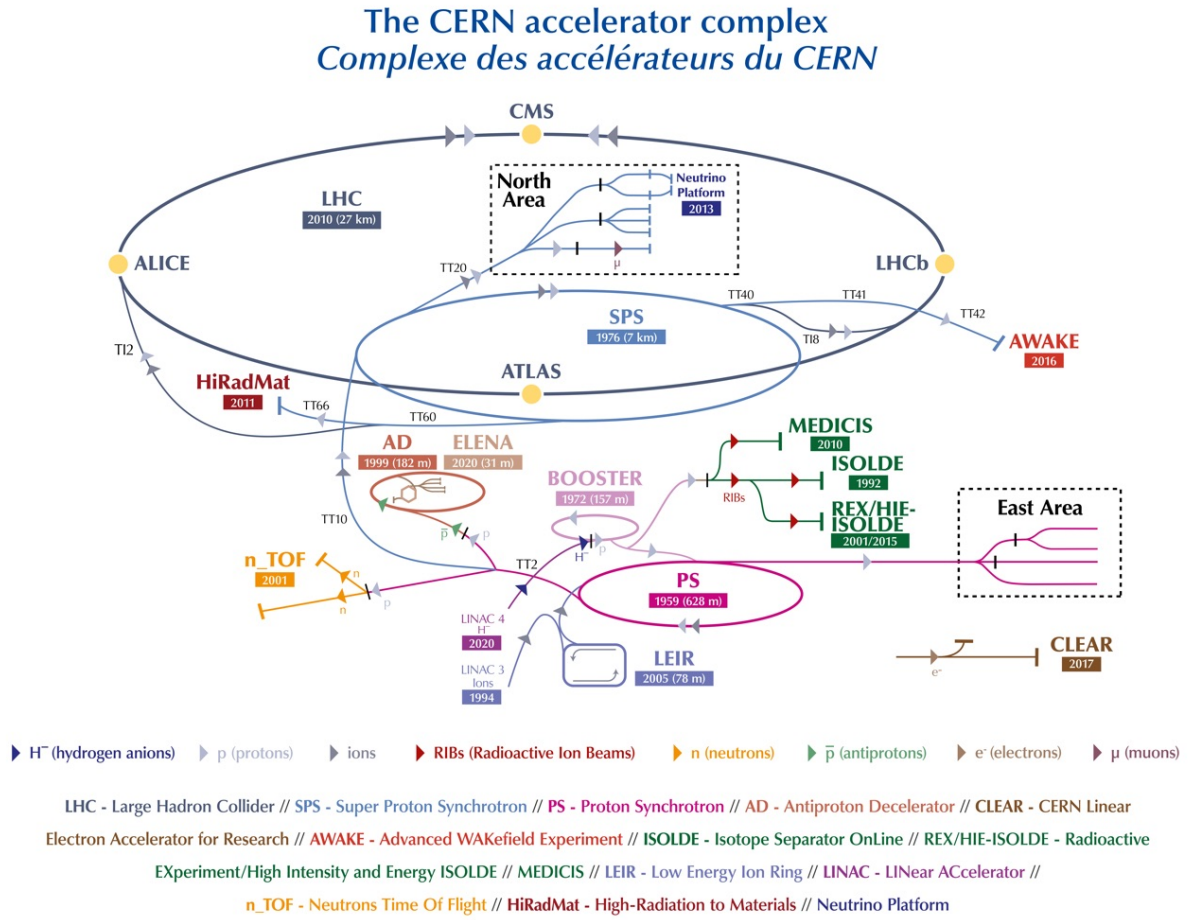


Figure 4.1: The CERN accelerator complex [45].

The CERN accelerator complex is a succession of different particle accelerators that brings particles to progressively higher and higher energy, ending with the LHC with a maximum per-proton-beam energy that is currently of 6.8 TeV.

The accelerator chain starts with the LINAC4, a linear accelerator of 86 meters that accelerates negative hydrogen ions up to 160 MeV. The ions then enter the Proton Synchrotron Booster (PSB) composed of four parallel synchrotron rings. In the PSB, they are stripped of their two electrons, leaving only protons that are accelerated up to 2 GeV for injection in the Proton Synchrotron (PS). The PS is a 628-meter ring that

accelerates protons from the PSB to 26 GeV. The protons then enter the last accelerator before the LHC, the Super Proton Synchrotron (SPS), a 6.9 km ring that brings their energy up to 450 GeV before LHC insertion. Protons from the SPS are transferred to the two beam pipes of the LHC, one beam circulating clockwise and the other anticlockwise, where they are both accelerated to the desired energy (multiple TeV).

As stated previously, the LHC itself is a 27 km long circular accelerator. It is composed of two beam pipes kept at ultra-high vacuum in which two proton beams circulate in opposite direction at near the speed of light. The accelerator is mainly composed of a large variety of magnets that allows to bend, focus and squeeze the proton beams. Several accelerating structures (electromagnetic resonators) are inserted at periodic interval to boost the energy of the protons along their trajectory. The magnets are made of super-conducting materials in order to obtain the large magnetic fields necessary to control such energetic proton beams. In particular, dipole magnets are used to bend the proton beams on the right orbit, quadrupoles allow to focus the beams along the way and more powerful quadrupoles are used to squeeze the beams when the two are intersecting in order to get the protons closer together and increase the chance of collisions. There are four interactions points (*IP*) at the LHC where beams intersect and collide. Around each IP a large particle detector is located. These four detectors are ATLAS, CMS, ALICE and LHCb. The LHC's main purpose is to provide stable beams with the highest possible energy and collision rate to the four experiments.

The LHC started up for the first time in September 2008, but due to technical problems the first actual data taking period (Run 1) started in 2010 with pp collisions at a centre-of-mass energy $\sqrt{s} = 7$ TeV. During the three years of Run 1, the LHC progressively increased the collision energy up to 8 TeV. Run 2 started in 2015 after a long shutdown where extensive repairs and upgrades to the LHC and the four experiments were conducted. The Run 2 collision energy was increased significantly to 13 TeV and lasted four years. After a second long shutdown, Run 3 started in 2022 at a record collision energy of 13.6 TeV. Run 3 is currently in progress and is planned to end in 2026. The data analysis presented in this thesis uses pp data from the two first years of the LHC's Run 3.

4.2 A complete description of the ATLAS detector

4.2.1 Overview

ATLAS (A Toroidal Large ApparatuS) is one of the four main experiments at the LHC located at the IP 1. As CMS, it is a general purpose particle detector conceived to be sensitive to the largest range of physics processes possible, in order to precisely study the SM of particle physics and search for BSM physics. ATLAS is able to identify and measure the properties (momentum, energy, mass, trajectory) of the particles produced during pp collisions. As shown in figure 4.2, the ATLAS experiment is a cylindrical detector of 44 meters in length and 25 meters in diameter with an overall weight of 7000 tons, making it the largest detector of the LHC. It surrounds the LHC beam line and is centred on the IP allowing ATLAS to cover nearly all the solid angle around the IP. Its internal structure is complex, composed of multiple cylindrical layers surrounding the beam line, each with a specific purpose.

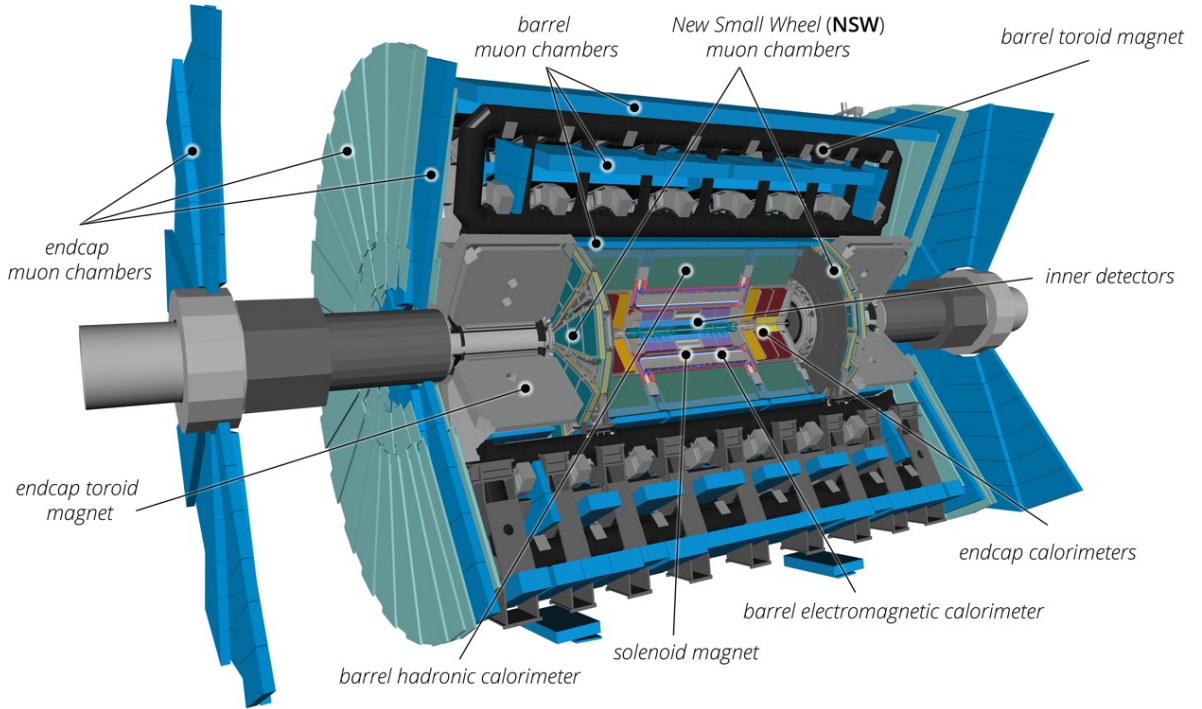


Figure 4.2: The ATLAS detector [46].

The ATLAS detector is composed of three large measuring instruments (*sub-detectors*), the inner tracker [47, 48], the calorimeters [49, 50] and the muon spectrometer [51], along with a system of magnets [52–55]. Generally, the sub-detectors are separated between a central (*barrel*) component and a frontward-backward (*end-caps*) component. In the context of this thesis work and the harnessed physics objects, a more detailed description of the inner detector and calorimeters is provided in this section compared to the muon spectrometer.

4.2.2 Coordinates system of ATLAS

Before getting to the description of the ATLAS sub-detectors it is important to define the system of coordinates used by ATLAS. ATLAS uses a cylindrical-like coordinates system which is represented in figure 4.3.

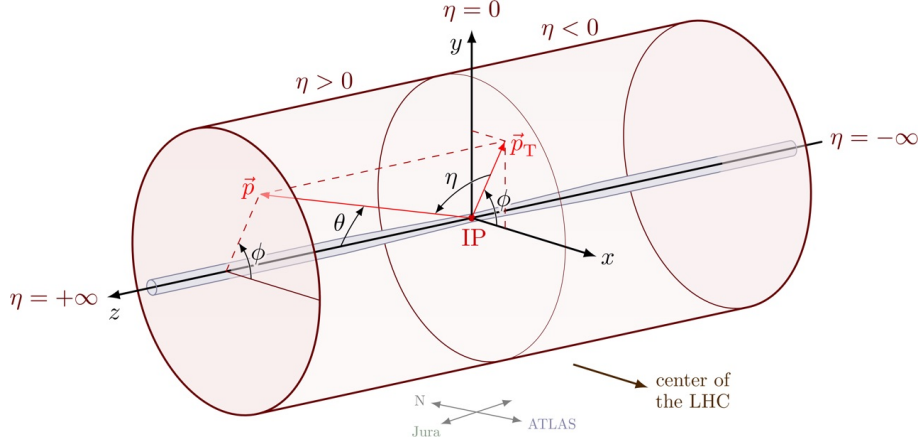


Figure 4.3: System of coordinates used by ATLAS [56].

The z-axis is aligned with the LHC beam line, the x-axis points to the centre of the LHC and the y-axis points upward. The origin of the coordinates is set to be the centre of the detector, at the IP. The x-y plane is the transverse plane to the beam line and so to the direction of displacement of protons. In this transverse plane the usual cylindrical coordinates are used (r, ϕ) where ϕ is the azimuthal angle around the z-axis. The final coordinate defining the angle regarding the z-axis (the beam line) differs from the usual cylindrical system. θ is not used directly but instead the rapidity or pseudo-rapidity. The rapidity y is defined as

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

where E is the energy and p_z the z-component of the momentum. This coordinate is defined such that it is a Lorentz invariant. y is equivalent to θ : $\theta = 0$ corresponds to $y = +\infty$, $\theta = \pi/2$ corresponds to $y = 0$ and $\theta = \pi$ corresponds to $y = -\infty$. Historically, the rapidity is not directly used but instead the pseudo-rapidity η defined as

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right)$$

which is an approximation of the rapidity, easier to compute. In the regime of low masses, the rapidity and pseudo-rapidity are similar: $\lim_{m \rightarrow 0} y \simeq \eta$.

To resume, positions in ATLAS are defined with this set of coordinates (r, ϕ, η) . The angular distance between two objects is vastly used in ATLAS and is defined as

$$\Delta R = \sqrt{\Delta \phi^2 + \Delta \eta^2} \quad (4.1)$$

Finally, to define the four-momentum $P = (E, \vec{p})$ ¹ of an object, the transverse momentum p_T is usually used. It is the momentum in the transverse plane:

$$p_T = \sqrt{p_x^2 + p_y^2} = |\vec{p}| \sin(\theta) = \frac{\sqrt{E^2 - m^2}}{\cosh(\eta)}$$

Such that an object is characterised by the set (E, p_T, m, η) .

4.2.3 Magnet system

Two systems of magnets allow to create an important magnetic field inside the detector. The solenoid magnet surrounds the inner tracker and produces a magnetic field of 2 T. The toroidal magnet produces a magnetic field of approximately 0.5 T for the muon spectrometer in the central region and of 1 T for the end-cap regions. Both aim at bending the trajectory of charged particles in order to get access to: the sign of the charge, given by the direction of curvature, and the momentum of the charged particles, given by the radius of curvature.

4.2.4 Inner tracker

The inner tracker (shown in figure 4.4) is the sub-detector closest to the beam line and provides tracking for the charged particles within $|\eta| < 2.5$. It is composed of three sub-detectors, the Pixel detector, the Semi-Conductor Tracker (SCT) and the Transition Radiation Tracker (TRT). Each is based on a different technology that allows to monitor the trajectory of charged particles produced during pp collisions, by measuring the crossing of these particles with different layers. Each measurement is called a hit and from a set of hits provided by each component of the inner tracker, the trajectory of a charged particle can be reconstructed (this will be detailed in 5.2). As this sub-detector is immersed in a large magnetic field, the curvature of the trajectory of a charged particle gives access to the momentum of the particle and the sign of its charge.

Figures 4.5 and 4.6 show respectively a transverse view and a longitudinal view of the inner tracker presenting the internal structures and positions of its different sub-detectors, which are described below.

The Pixel detector The Pixel detector is the first sub-detector of the inner tracker crossed by charged particles produced at the IP. It is composed of four cylindrical layers² wrapped around the beam line in the barrel region and three disks perpendicular to the beam line in each end-cap region. The three outer layers and disks are made of silicon pixel sensors with an $r\phi$ (transverse) and z (longitudinal) segmentation and size of $50 \times 400 \mu\text{m}$. The innermost layer is composed of smaller pixel sensors with a size of $50 \times 250 \mu\text{m}$. The four cylindrical layers are positioned respectively at radii 33.25, 50.5, 88.5 and 122.5 mm with a coverage up to $|\eta| = 2.5$ and with an accuracy of $10 \mu\text{m}$ in $r\phi$ and $115 \mu\text{m}$ in z . The disks are placed along the beam line at $|z|$ of 495, 580, 650 mm with an accuracy of $10 \mu\text{m}$ in $r\phi$ and $115 \mu\text{m}$ in r . In total the inner tracker has 92 million readout pixel channels. Typically, charged particles produced at the IP will cross the four barrel layers leaving four measured hits for trajectory reconstruction.

¹ Where E is the energy, $\vec{p}$ the momentum and the mass is given by $m^2 = P^2 = E^2 - \vec{p}^2$

² The innermost one is the Insertable Barrel Layer (IBL) and was installed during the first long shutdown

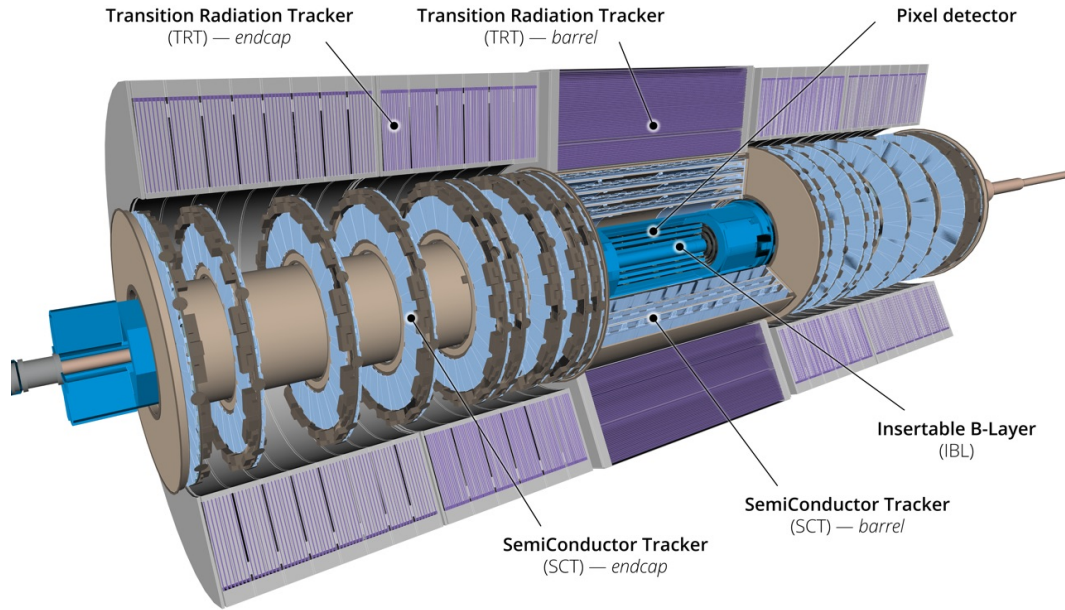


Figure 4.4: Overall view of the inner tracker of ATLAS [46].

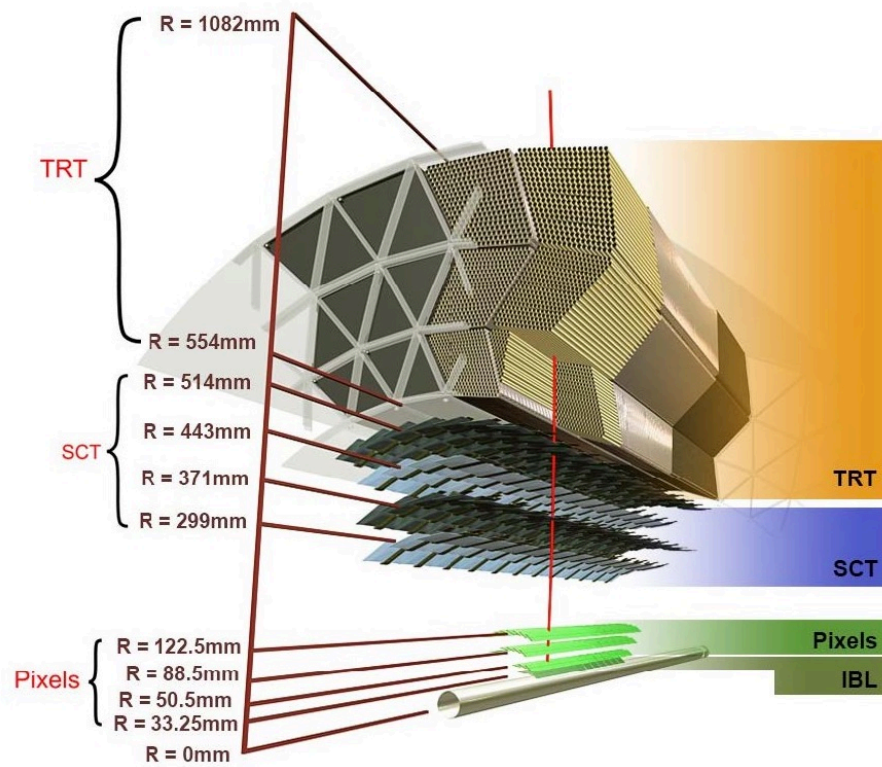


Figure 4.5: Transverse view of the inner tracker [46].

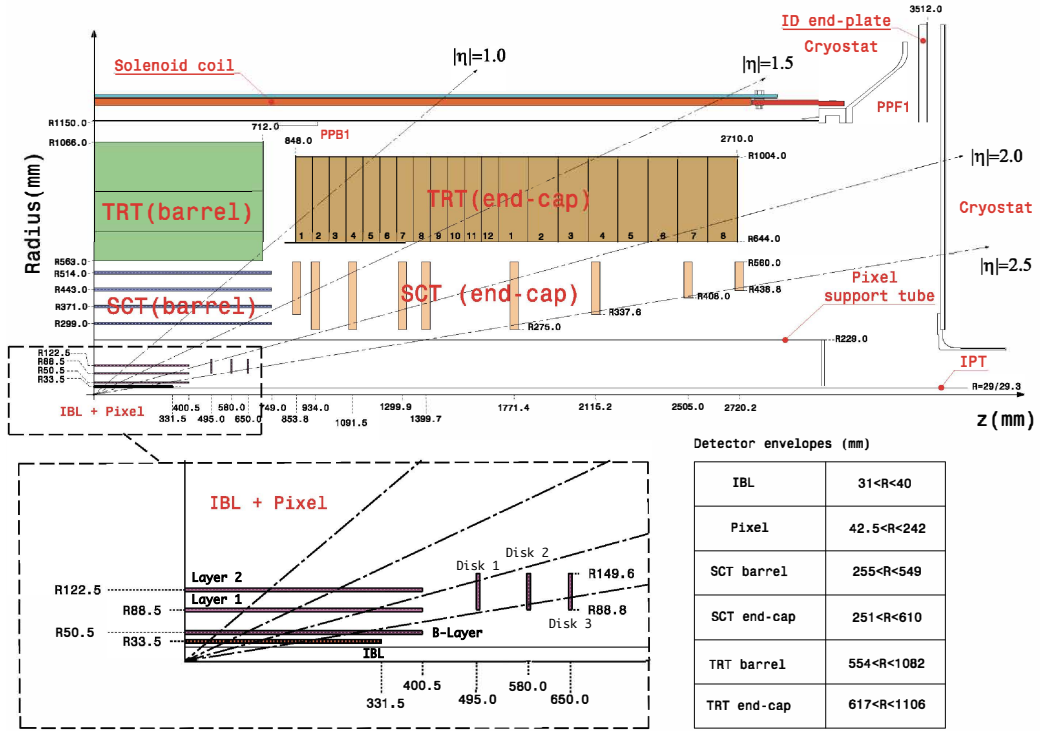


Figure 4.6: Longitudinal schematic of the inner tracker [46].

The Semi-Conductor Tracker Following the Pixel detector is the SCT. It is also composed of cylindrical layers in the barrel region and perpendicular disks for the end-cap regions covering a surface of 63 m² with silicon modules and providing an almost hermetic coverage. There are four layers using stereo strips to measure both $r\phi$ and z , including one set of strips that is parallel to the beam line measuring $r\phi$. They are positioned at radii 299, 371, 443 and 514 mm with a coverage up to around $|\eta| = 1.5$ and an accuracy of 17 μm in $r\phi$ and 580 μm in z . The perpendicular disks use a set of strips running radially measuring r and a set of stereo strips measuring ϕ . The disks extend the coverage of the SCT up to $|\eta| = 2.5$ and have an accuracy of 17 μm in $r\phi$ and 580 μm in r . There are 9 disks on each side of the detector. The SCT counts 6.3 million readout channels and typically provides 4 measured hits for each charged particle.

The Transition Radiation Tracker The last and outermost sub-detector of the inner tracker is the TRT. It is made of a succession of gas-filled (Xe/CO₂/O₂ mixture) straw tubes interleaved with transition radiation material extending from a radius of 554 mm up to 1082 mm. In the barrel region, the tubes are parallel to the beam line and work as drift tubes and thus only provides a measurement of r and ϕ with an accuracy of 130 μm , with an $|\eta|$ coverage up to around 1.0. In the end-cap regions, the tubes are perpendicular to the beam line and arranged radially in wheels providing z and ϕ measurements with the same accuracy as in the barrel region. The end-cap components of the TRT extend the η coverage around 2.0. The TRT totalise 351,000 readout channels and provides typically 36 hits per charged particles.

The combination of these three sub-detectors gives several dozens of high precision hits and allows for a robust pattern recognition and trajectory reconstruction. From the measurements of the inner tracker, reconstructed trajectories called *tracks* are obtained.

4.2.5 Calorimeter system

The calorimeter system of ATLAS is built around the inner tracker and handles the measurements of the energy and direction of particles travelling through it. It is designed to make the particles interact electromagnetically or strongly through the development of calorimeter showers, making them deposit all their energy. Figure 4.7 presents the arrangement of the ATLAS calorimeter system. This system can be separated in two main sub-systems: the closest to the beam line is the electromagnetic calorimeter (EMCal) which measures and stops incoming photons and electrons³, and surrounding it is the hadronic calorimeter (HCal) which measures and stops hadrons. As for the inner tracker, each of the EMCal and HCal has a barrel and end-cap components for a coverage up to $|\eta| = 3.2$. In addition, forward calorimeters are positioned very close to the beam line on each side of the detector extending the calorimeter coverage up to $|\eta| = 4.9$ and allowing to have a near complete coverage of the solid angle around the IP.

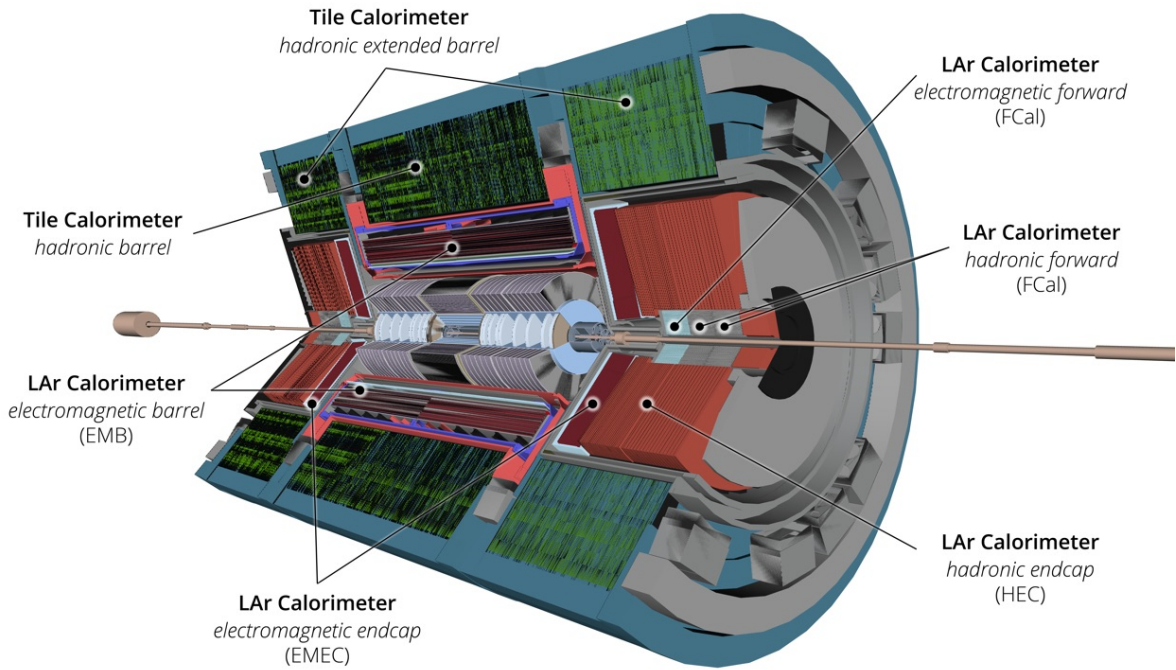


Figure 4.7: Internal view of the calorimeter system of ATLAS [46].

In ATLAS, the energy measurement is based on a sampling technology: particles go through a succession of high-density materials (absorber) that stop the particles and active materials that measure the particles energies. A particle crossing a calorimeter will interact electromagnetically or strongly (depending on the type of particle or sampling technology) with the high-density materials producing an electromagnetic or hadronic shower of secondary particles. These secondary particles then travel through the active layers where their energies are actually measured. This phenomenon continues until the energy of the incoming particle is completely absorbed by the high-density materials. The total energy measurement of the particle is then given by the sum of the energy measurements of all the secondary particles forming the shower.

For ATLAS, there are two sampling calorimeter technologies: liquid argon (LAr) is used as the active

³ In the text, no distinction is made between electrons and positrons, the term electron is used to enclose both particles

medium for the EMCal (with lead absorber), the end-cap (with copper absorber) and forward (with copper + tungstene absorber) calorimeters; and in the barrel region of the HCal the active medium used are plastic scintillating tiles while the absorber is made of steel. Each calorimeter is segmented in cells in the $\eta - \phi$ plane with a finite granularity, allowing to measure the longitudinal and azimuthal position of energy deposits. Calorimeters in the barrel region are also segmented in several cylindrical layers around the beam axis providing an additional depth (transversal) information on the localisation of the energy deposits. The main characteristics of the different calorimeters are given in Table 4.8.

Figure 4.8: Summary table of the main parameters of the sub-detectors of the ATLAS calorimeter system [46].

	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	$1.425 < \eta < 2.5$
Calorimeter 3rd layer			0.1×0.1	$2.5 < \eta < 3.2$
	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)	

4.2.6 Muon spectrometer

The muon spectrometer forms the outer layers of the ATLAS detector. As the name indicates, it aims at measuring the trajectories of muons and determine their momentum (like the inner tracker for generic charged particles). The muons are almost the only particles, with the neutrinos, that can cross the full thickness of the calorimeters and escape the detector. This is due to its large mass leading to weak bremsstrahlung radiation. The muons leaving the calorimeters enter the muon spectrometer where a magnetic field bend their trajectory which is measured by multiple layers of Monitored Drift Tubes (MDT) and Resistive Plate Chambers (RPC) in the barrel region and by Thin Gap Chambers (TGC) and MDT in the frontward regions. Figure 4.9 shows the layout of the muon spectrometer.

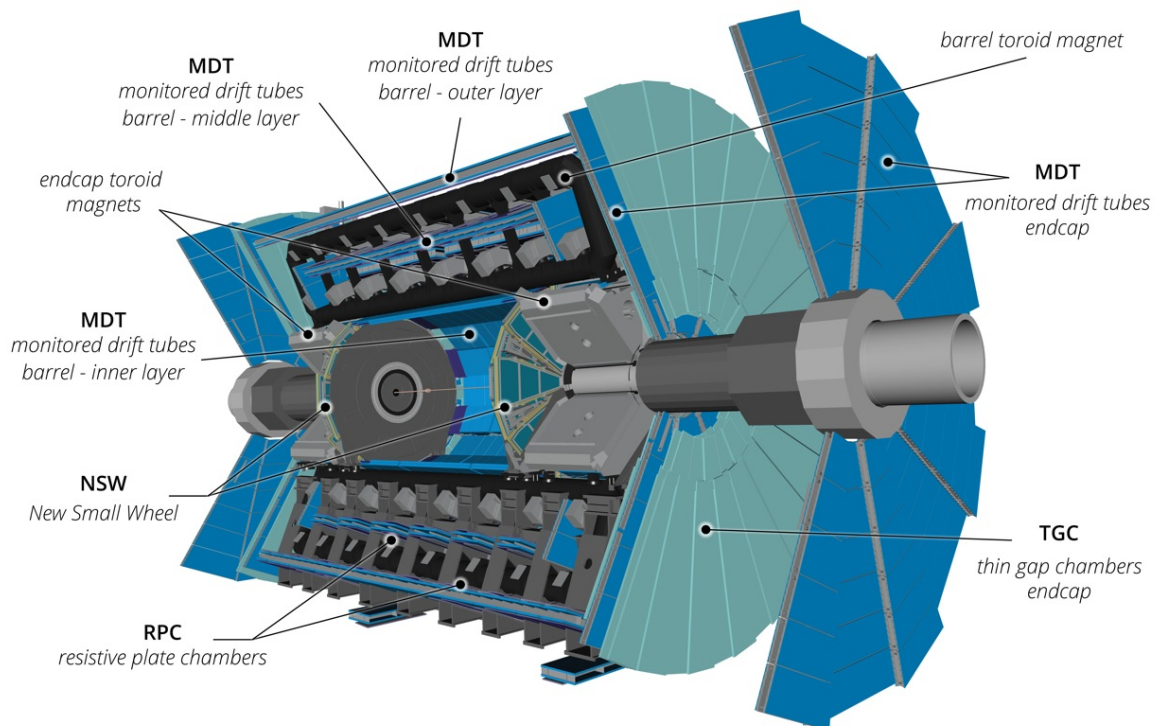


Figure 4.9: The muon system [46].

4.3 Data taking with ATLAS

The LHC provides pp collisions to the ATLAS experiment. A collision and the particles it produces is called an *event*. The quantity, rate and characteristics of the events provided by the LHC depends on the proton beams conditions. Which events are recorded by ATLAS on the other hand depends on its trigger system.

4.3.1 LHC conditions: luminosity and pileup

As mentioned in Section 4.1, one of the performance parameter of the LHC is its collision rate and more precisely its luminosity. The luminosity is the measure of the rate of produced events of a particular process characterised by a cross-section σ . It measures the ability of the LHC to produce a specific number of events per processes. The general definition of the luminosity $\mathcal{L}$ is the proportional factor between the cross-section σ and the rate of produced events $\frac{dN_{\text{evt}}}{dt}$:

$$\mathcal{L} \times \sigma = \frac{dN_{\text{evt}}}{dt} \rightarrow \mathcal{L} = \frac{1}{\sigma} \frac{dN_{\text{evt}}}{dt} \quad [57]$$

And is expressed in $cm^{-2}.s^{-1}$. Maximising the luminosity is then crucial to produce the maximum number of events possible for each process. It is especially important for rare processes. The luminosity at the LHC depends on the characteristics of the proton beams. Along the CERN accelerators complex, the protons are manipulated to create a specific beam structure. In these beams, the protons are regrouped in bunches of N particles that are evenly distributed along the LHC ring with about 25 ns separation. During Run 3, the beams are made of about 2800 bunches each containing around 2.10^{11} protons. The luminosity for this particular setup is then defined as:

$$\mathcal{L} = \frac{N^2 f N_b}{4\pi\sigma_x\sigma_y} \quad [57]$$

Where:

- N is the number of protons per bunch
- f is the revolution frequency of the bunches
- N_b is the number of bunches in each beam
- σ_x and σ_y are the bunch sizes in the transverse plane at the interaction points

The luminosity during Run 1 and Run 2 was progressively increased by changing the characteristics of the beams to obtain a peak luminosity of $2.10^{34} cm^{-2}.s^{-1}$ obtained in 2018, twice the design luminosity. For Run 3, accelerator improvements allowed sustained running at this luminosity.

Another quantity to measure the performance of the LHC and which is directly related to the luminosity, is the integrated luminosity $\mathcal{L}_{\text{int}}$. It is simply the luminosity integrated over time:

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

Usually expressed in invert femtobarns fb^{-1} ($1\text{b} = 1\text{cm}^{-2}$) which is the right unit to effectively describe what the LHC provides to the experiments. This is a quantity of interest as it relates directly to the number of events produced for a certain process:

$$\mathcal{L}_{\text{int}} \cdot \sigma = N_{\text{evt}}$$

Figure 4.10 presents the evolution of the integrated luminosity delivered to ATLAS by the LHC over the years during Run 1 and Run 2. In total, during Run 1, 28 fb^{-1} at $\sqrt{s} = 7 - 8 \text{ TeV}$ was delivered to ATLAS and 157 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$ during Run 2. For Run 3, the LHC is running at a record centre-of-mass energy of $\sqrt{s} = 13.6 \text{ TeV}$ with already 70 fb^{-1} delivered to ATLAS (and the other experiments) in 2022 and 2023 and is expected to provide more than 250 fb^{-1} in total.

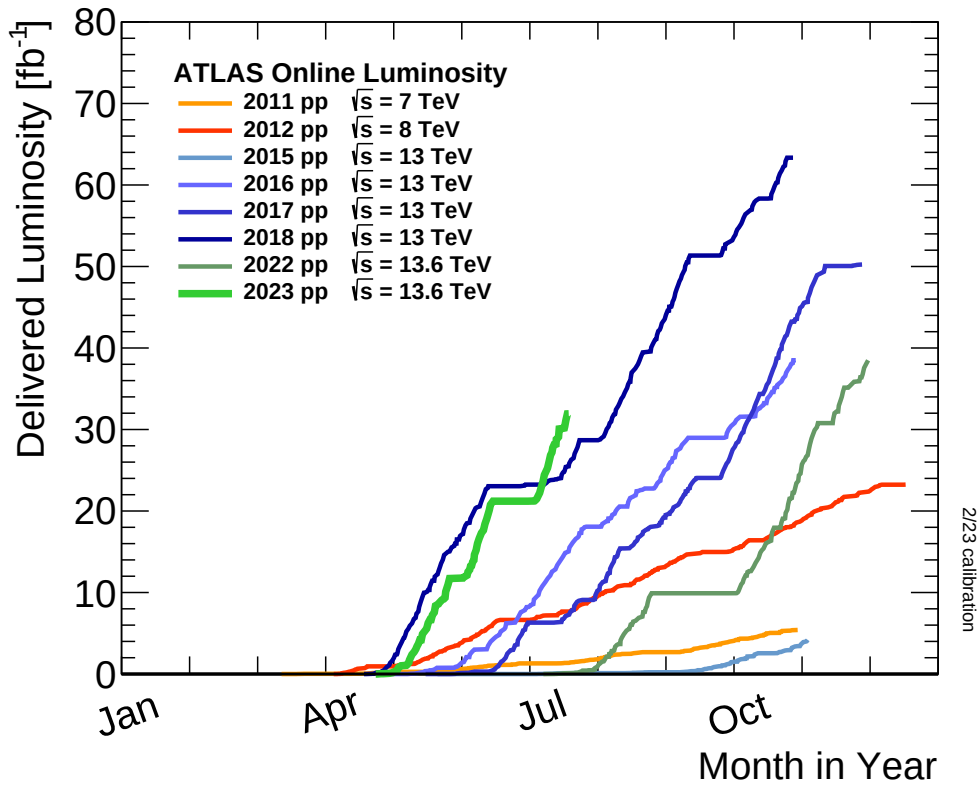


Figure 4.10: Integrated luminosity over the years delivered by the LHC to ATLAS during Run 1, Run 2 and the beginning of Run 3 [58].

However, there is a difference between what the LHC delivers and what ATLAS exploits for physics analyses. First, because of inefficiencies like dead-time or machine commissioning periods, the integrated luminosity recorded by ATLAS is lower than what the LHC deliver. Secondly, not all events recorded are certified for physics analysis usage as anomalous conditions can compromise the data (like detector failures). Figure 4.11 (a) shows during Run 2 the total integrated luminosity: delivered by the LHC, recorded by ATLAS and good for physics. For Run 2, the total efficiency of data taking good for physics is of about 95% of the total periods where the LHC provided stable beams [59]. Figure 4.11 (b) shows also the first delivered and recorded integrated luminosity for the beginning of Run 3.

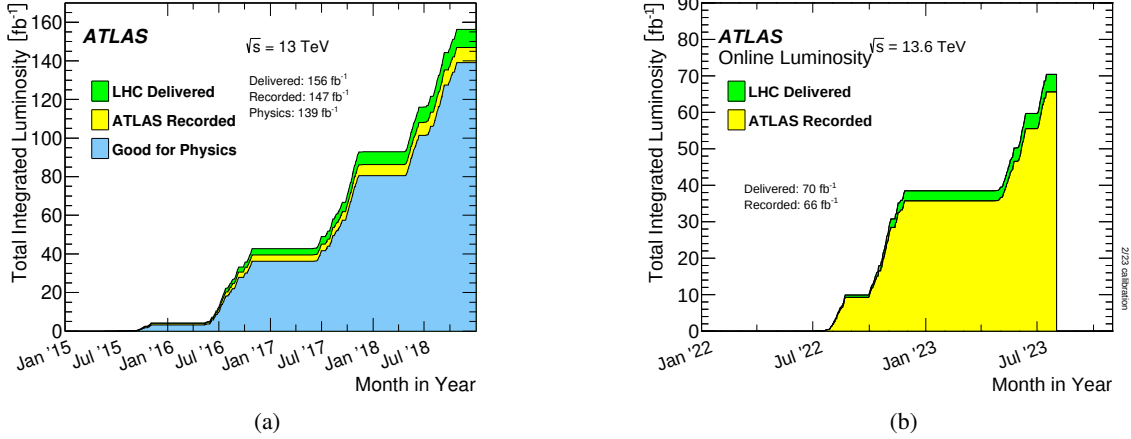


Figure 4.11: Integrated luminosity over the years delivered by the LHC (green), recorded by ATLAS (yellow) and good for physics (blue) during Run 2 (a) and beginning of Run 3 (b) [58].

At the same time, with great luminosity comes great difficulties in data taking for all experiments. At LHC high luminosity is achieved thanks to reduced time separation between bunch crossings (about 25 ns) and high density of protons (about $2 \cdot 10^{11}$ protons per beam and small beam sizes). This leads to two problematic phenomena. First, more than one collision happen in the same bunch crossing and so events are polluted by neighbouring collisions. Secondly, the sub-detectors time response can be greater than the time between two bunch crossings (it is especially the case for the LAr calorimeters) such that the previous and next bunch crossings can, again, pollute the measurements of an event. In the first case, this phenomenon is called *in-time pileup* and in the second case *out-of-time pileup*. If no distinction is made, the term *pileup* is generally used and includes both types. It is usual to differentiate the hard-scatter collisions that produce events of interest for physics analyses and the ones that pollutes the said events which are called pileup events.

Pile-up is a great difficulty for analyses because of the necessity to identify the particles coming from the events of interest in order to only consider these particles in the analyses. The subtraction of particles coming from pileup events is a difficult task and is generally not perfectly possible. Charged particles from pileup that produce tracks in the inner detector can be removed as they can be associated to a different collision to the one identified as the origin of the event of interest. But neutral particles are much more difficult to remove individually as they deposit energy in calorimeters along the particles from the events of interest. Out-of-time pileup can be mitigated by cutting on the time of arrival of the energy deposits. However, for in-time pileup only an average subtraction over all the calorimeter area is possible as it is not possible to distinguish the deposits. This subtraction is based on an estimate of the in-time pileup contribution to the energy deposits, which is itself based on the pileup rate (this will be detailed in Section 5.4).

The measurement of the pileup rate is then an important task. Different quantities are used, the estimation of the actual or mean number of interactions/collisions per bunch crossing (respectively noted μ and $\langle \mu \rangle$) and the number of reconstructed collisions (noted N_{PV} for Number of Primary Vertex⁴) are the main ones. Figure 4.12 shows the recorded integrated luminosity of ATLAS as a function of the mean number of interactions per bunch crossing during Run 2. It shows that $\langle \mu \rangle$ varied during Run 2 from 10 to 70

⁴ The identification of primary vertices will be described in Section 5.3

interactions per crossing with an average of 33.7. During the two first years of Run 3, on average a $\langle \mu \rangle$ of 60 interactions per bunch crossing was obtained. The higher these quantities are (μ , $\langle \mu \rangle$ or N_{PV}), higher the pileup rate is and so the quantity of pileup particles.

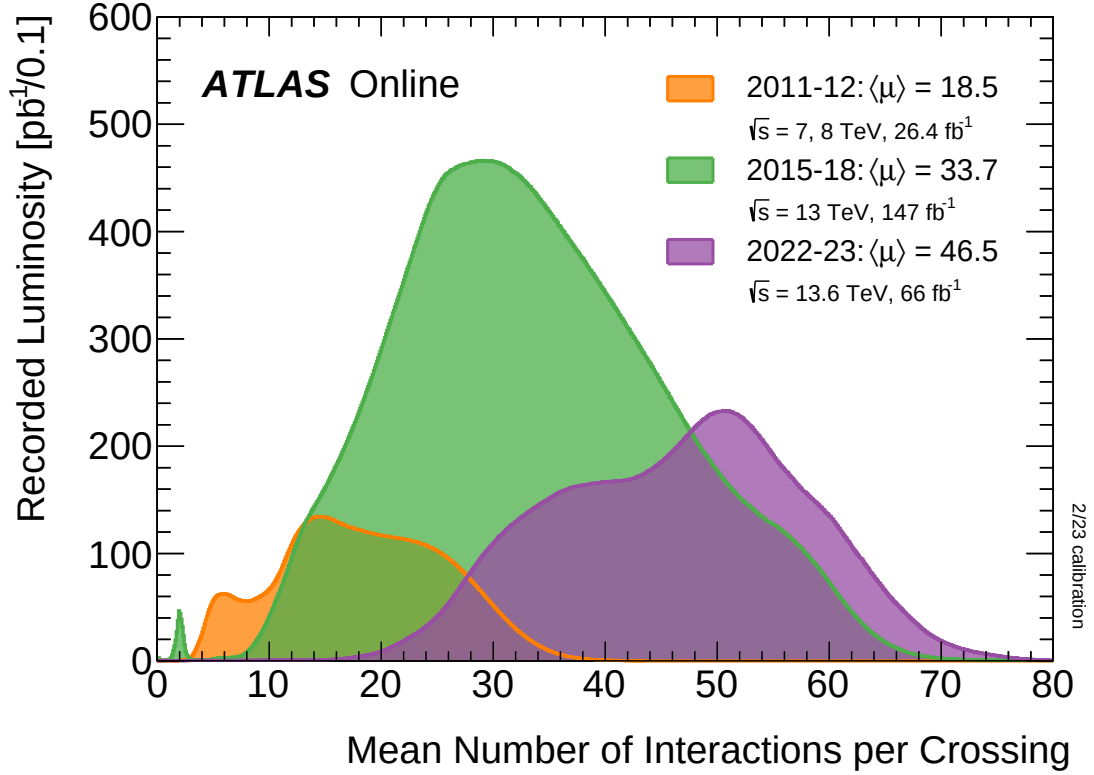


Figure 4.12: Recorded integrated luminosity as a function of the mean number of interactions per bunch crossing with ATLAS during Run 1 (orange), Run 2 (green) and beginning of Run 3 (purple) [58].

4.3.2 Trigger system

In ATLAS, there is a bunch crossing every 25 ns, meaning at a rate of 40 MHz, and the size of an event is approximately 3 Mb (sum of all the readout channels responses in the different sub-detectors). It is thus physically impossible to record every event (as it would correspond to a data rate of $\sim 120 \text{ Tb/s}$). The rate must be decreased, this is done thanks to a trigger system. The trigger system selects, based on physics criterion, the events to actually record in order to keep only the events with a physical interest for analyses⁵. This system is composed of two sub-systems, the Level 1 trigger (L1) and the High Level Trigger (HLT). L1 is a hardware-based trigger, whereas the HLT is a software-based trigger.

The L1 trigger is based on custom electronics and do not rely on software computation, making it really fast. All the event processing must be done within the time latency allowed by the detector electronics, this latency is $2.5 \mu\text{s}$ per event. It uses reduced granularity information from the calorimeters and muon spectrometer to search for events with high-energy signatures, like high-energy electrons, muons, photons or jets, and events with large missing energy in the transverse plane. L1 allows also to determine region of

⁵ A large proportion of the collisions are low energy scattering processes and are thus not interesting for physics.

interests (RoI) in the detector in order to reduce the search regions for the HLT. These RoI are regions where L1 spotted high-energy deposits in the calorimeters and/or tracks in the muon spectrometer. The L1 trigger reduces the rate from 40 MHz (the LHC collision rate) to 100 kHz (the L1 acceptance rate).

The events accepted by L1 are then processed by the HLT, which is a much higher-level trigger based on software computation, able to reproduce offline (i.e. analyses level) selections as closely as possible. It is slower than L1 with on average a per event processing time of 400 ms. It starts from RoIs defined by L1 to reconstruct events at higher levels of details than L1. It uses the full granularity of the calorimeters, muon spectrometer and tracking system to reconstruct high-level physics objects (tracks, jets, electrons, photons, muons, MET, HT, . . .). The HLT selection criteria are more perceptive than for L1, they can span from usual selections like high-energy electrons or high-energy jets to more complex signatures like emerging jets. The criteria are chosen to maximise the data available for important analyses and are defined in a trigger menu that is often updated to adapt to new analyses needs and ideas. At the end of the HLT, the rate is further reduced, down to 3 kHz, making it much more acceptable for data storing.

4.3.3 Data taking workflow

The events that are selected by the trigger system are then sent to be stored in data centres. There, from the raw detector measurements they are fully reconstructed using offline software (this will be detailed in Chapter 5) to produce high-level physics objects, like tracks, vertices, jets, electrons, photons and muons, with higher precision and details than at the trigger level. Once this is done, they are made available for physics analyses in specific datasets.

4.4 Monte Carlo simulation

Monte Carlo (*MC*) simulations reproduce pp collisions and the ATLAS detector response. They are used for both physics analyses and studies of the detector performance. In physics analyses, MC simulation can be used to study theoretical signals (like SUSY or emerging jets), estimate background or design selection criterion. Studies of the detector performance possible with MC simulation also allow to derive calibration for physics objects. MC simulations are made of two distinct parts, the particle generator and the detector response simulation. The former simulates the particle production in the collisions and the subsequent dynamics of the SM or of the BSM theoretical models, while the later simulates the interactions of the particles produced by the generator with the detector materials.

The main examples of particle generators commonly used in particle physics are PYTHIA [60], SHERPA [61], HERWIG [62] and MADGRAPH [63]. They are able to simulate pp collisions, taking into account the parton distribution function in the protons, similar to the LHC conditions. They support most of the SM interactions accessible from the collision of protons and simulates all the dynamics to pass from the emission of the products of the interaction to the final stable particles. It includes the simulation of QCD parton shower and hadronisation following the emission of high-energy quarks and gluons, the emission of radiations before and after the collision and the decay chains of hadrons. The product of the particle generator is a list of stable particles described by their production vertices and four-momentums. In ATLAS the particle generators are generally used only to simulate one pp interaction per event, the one leading to events of interest (e.g. hard-scatter interaction producing highly energetic quarks, gluons or bosons). To be in agreement with the LHC conditions, pileup interactions are added by superposition to the main collision event. These pileup interactions are generated separately and are designed to correspond to the pileup rate observed in the real data. As the conditions of the LHC can vary during a Run, different pileup configurations are generated, each one corresponding to a certain LHC condition and are used to create different MC datasets.

Following the particle generation is the detector simulation. ATLAS uses the software GEANT4 [64, 65] to simulate the detector response to the generated particles. GEANT4 uses a very detailed modeling of ATLAS. This model includes both the active materials which provide measurements and dead materials including internal structure supporting the detector components and service cables⁶. The dead materials do not participate in the measurements themselves but in the overall detector response as they can absorb a part of the particles energy. With this ATLAS model and for each generated particle, GEANT4 simulates step by step the trajectory of the particles and their potential interactions with any active or dead materials. Finally, all the energy deposits in the inner detector, calorimeters and muon spectrometer are converted in electronic signals as they would have been measured by the real detector, this step is called *digitalisation*. After this point, the MC simulated events are computationally identical to raw measurements from the real detector. The same reconstruction software applied to real data can then be used on the simulated events to reconstruct physics objects.

In the end, the real data and the MC simulation are identical in terms of reconstructed objects. However, the MC simulation have the particularity to contain information at the particle level before the detector response (called *truth* level), it allows comparing truth to reconstructed objects which is crucial for understanding and studying the detector response.

⁶ Which provide current and cooling to the sub-detectors and retrieve measured signals from the readout channels

CHAPTER 5

RECONSTRUCTION OF COLLISION EVENTS AND PHYSICAL OBJECTS WITH THE ATLAS DETECTOR

Outline of This Chapter

5.1	Reconstruction workflow, from raw detector measurements to complete physical reconstruction	45
5.2	Tracks: charged particles trajectory reconstruction	47
5.3	Particle decay vertex reconstruction	51
5.4	Jet reconstruction	55

The events recorded by the ATLAS detector or produced by MC simulations are grouped and stored in datasets in which each entry corresponds to an event and contains the measurements from every readout channels of the detector. These measurements are the hits in the inner tracker and the muon spectrometer and the energy in each cell of the calorimeters. A complex chain of reconstruction is applied to these raw datasets to obtain datasets exploitable for analyses where each entry is a reconstructed event, containing reconstructed physics objects (tracks, jets, muons, . . .) and event metadata. This reconstruction chain is performed using a python-C++ framework called *Athena* [66].

In this chapter, an opening section describes the reconstruction workflow undergone by the raw datasets (5.1), then Sections 5.2, 5.3 and 5.4 delve into the reconstruction techniques and algorithms for tracks, vertices and jets respectively. Only details on these objects are given here as they are the main objects used in the rest of this thesis. A particularly extensive description of the jet reconstruction is presented as a dedicated project of my thesis is focused on jet calibration.

5.1 Reconstruction workflow, from raw detector measurements to complete physical reconstruction

After both real data recording by the ATLAS detector and MC simulation digitalisation step, raw datasets are obtained containing hit information from the inner detector and the muon spectrometer (position coordinates), along with information from the calorimeter cells (energy and position measurements). This file format is called RDO for *Raw Data Outputs* and is the initial format for event reconstruction. From this file format, a first processing step, called reconstruction, is applied to build basic objects for analysis use. In this step, among others, tracks are reconstructed from the hits in the inner detector, and energy clusters are formed from the calorimeter cells' information. Reconstruction allows a first reduction in the quantity of information, saved as higher-level reconstructed objects. This new file format is called AOD for *Analysis Object Data*. Then, a second processing step, called derivation, is applied to produce DAOD files (*Derived Analysis Object Data*). This is a step where various analysis needs are taken into account for the reconstruction of high-level physics objects and the selection of the needed information. During this step:

- vertices and jets are reconstructed,
- object and variable selections (*thinning*) are applied to only keep what is necessary for the analyses¹,
- an event-level selection procedure (*skimming*) is applied which only saves events firing the desired triggers.

Derivation allows for a second reduction in the size of datasets by removing information but also entire unwanted events. The obtained DAOD datasets are thereby ready for analyses manipulations. This workflow is resumed in figure 5.1.

¹ Objects with high enough energy, tracks linked to other important physics objects like high-energy jets or reconstructed vertices, ...

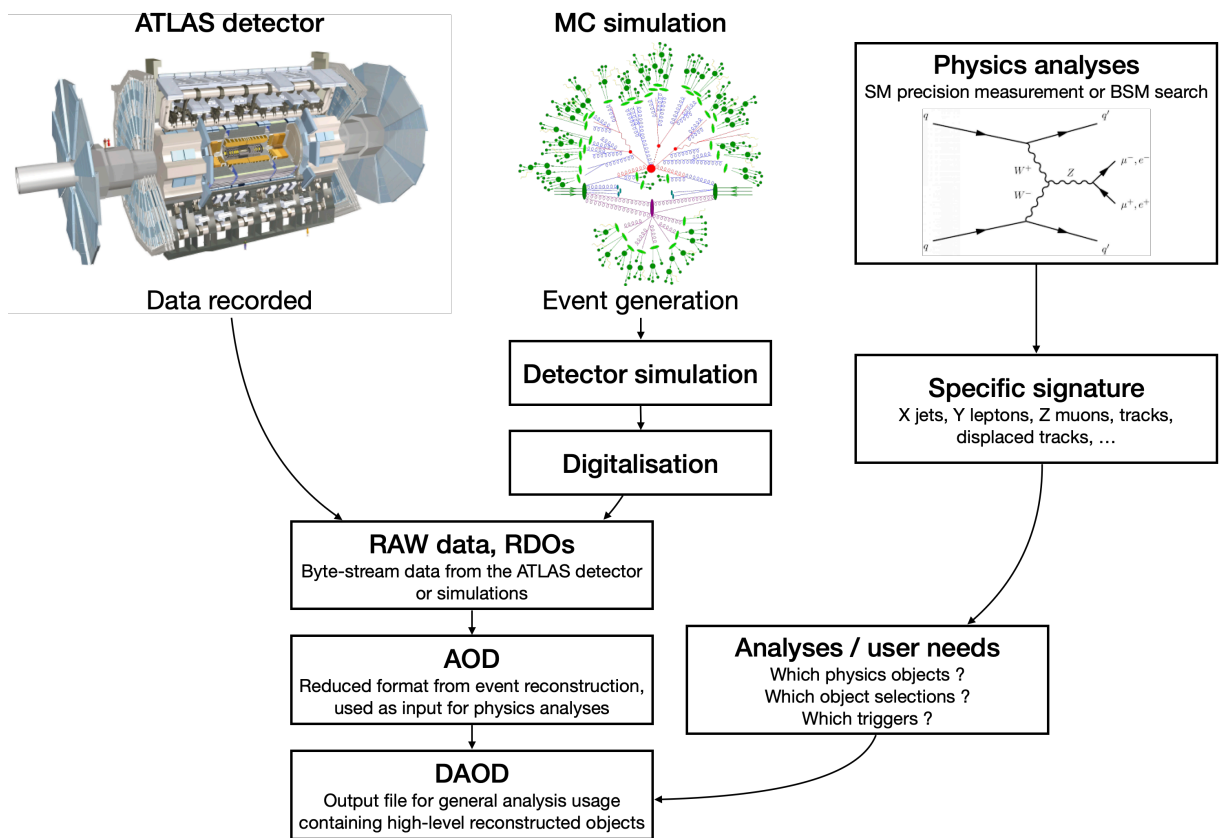


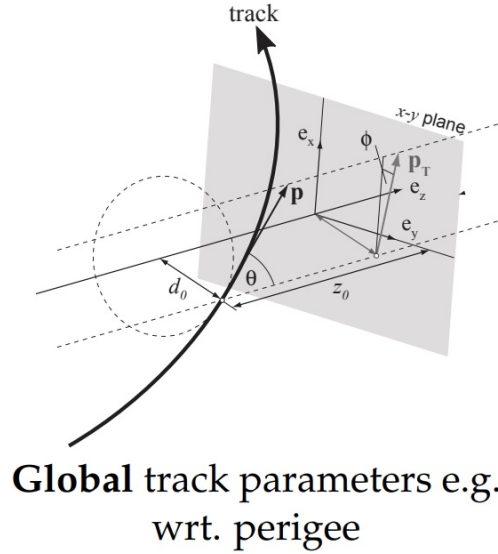
Figure 5.1: Reconstruction workflow of real data recorded by ATLAS and MC simulation to DAOD.

5.2 Tracks: charged particles trajectory reconstruction

Charged particles trajectories (*tracks*) in a magnetic field form segments of helices that can be described by five parameters:

$$(l_x, l_y, \phi, \theta, Q/P)$$

l_x and l_y are local coordinates of a reference point in the particle trajectory, ϕ and θ define the direction of propagation of the particle at the reference point, and Q/P is the ratio between the charge of the particle and its momentum which corresponds to the track curvature. These parameters depend on a chosen reference axis and point. In ATLAS, the beam-line (z-axis) is aligned with the magnetic field in the inner detector and so is the default reference axis for tracks with the origin set at the beam spot (mean position of the pp collisions). The reference point for the local coordinates is chosen to be the single point of closest approach to the beam spot: l_x is then noted d_0 and gives the distance of the reference point to the beam spot in the transverse plane, and l_y is noted z_0 and gives the z-coordinate of the reference point to the beam spot. This representation is called the *perigee* representation and is schematized in Figure 5.2. (d_0, z_0) are called respectively transverse and longitudinal impact parameters. Both give a proxy of the position of production of the tracks: the further away of the beam spot the track is produced, the greater its impact parameters will be. High (d_0, z_0) tracks are indicators of displaced production of particles², it is hence crucial for long-lived particles (LLP) analyses, as the one that is presented in this thesis, to reconstruct charged particles trajectories as robustly and precisely as possible.



$$\left(d_0, z_0, \phi, \theta, \frac{q}{p} \right)$$

Figure 5.2: Schematic of the track parameters convention used by ATLAS [67].

² Prompt particles production are assumed to happen close to the beam spot.

Track reconstruction in ATLAS aims at combining hits from the inner tracker deposited by the same charged particle and estimating the corresponding track parameters:

- d_0, z_0 , the impact parameters,
- ϕ, θ or more generally η , the direction of propagation of the track,
- the transverse momentum p_T of the track computed from the track curvature Q/P .

It is a challenging task due to the combinatorial complexity of several thousands of hits per event in the inner detector that need to be processed and combined. There are three stages in the ATLAS track reconstruction software [68, 69]. Two are resumed in Figure 5.3 that presents the two main, consecutive, reconstruction chains: an inside-out stage to find tracks originating from the beam line³ with a scan from the inside to the outside of the inner detector, and a back-tracking stage, scanning left-over hits from the outside to the inside of the inner tracker. The last stage, called Large Radius Tracking (*LRT*), aims at reconstructing tracks produced further away from the IP. These stages output two collections of reconstructed tracks, one containing prompt tracks originating from the beam line and its vicinity (primary + secondary chains) and a second one containing large radius tracks which do not originate from the primary pp collisions (*LRT*).

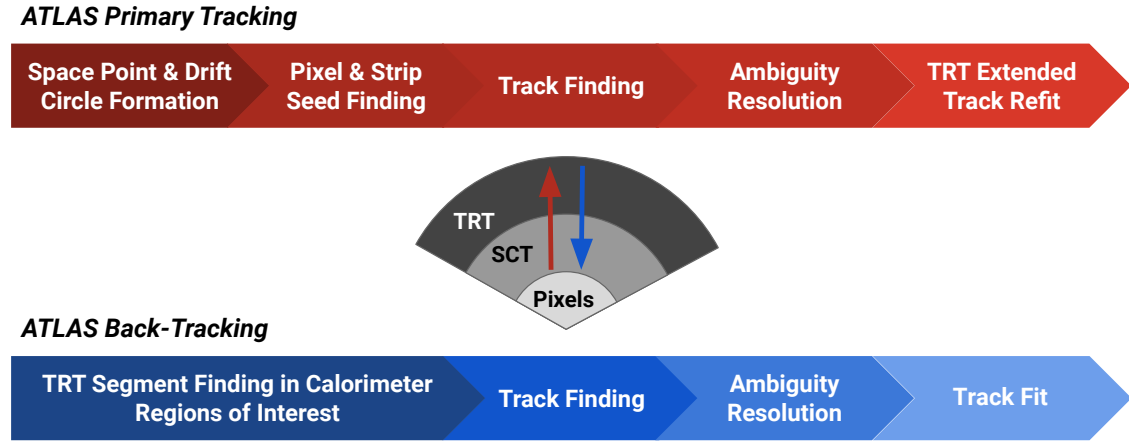


Figure 5.3: Schematic view of the primary and secondary tracking chains used in ATLAS inner detector track reconstruction [68].

The track reconstruction procedure starts with a pre-processing of the read-out channels from the different sub-detectors of the inner tracker. The energy deposits of incoming charged particles are converted into three-dimensional points (the hits) for the Pixel and SCT sub-detector and two-dimensional hits for the TRT.

Primary tracking chain, inside-out pass The first stage of track reconstruction is the inside-out pass, it aims at identifying tracks originating from the primary pp collision, close to the beam line, with small d_0 impact parameter. It starts by forming track seeds from the association of three hits in either the Pixel and SCT sub-detectors or a combination of the two. The three hits are associated based on their compatibility with them originating from the same track. These seeds are then filtered on some loose selection criteria to

³ With small impact parameters.

remove the ones that would produce low-quality tracks and to improve the computation time. From the selected seeds, search roads are constructed from the remaining Pixel and SCT hits that are compatible with the same seed and its estimated trajectory in the tracker. This is to reduce combinatorics by only searching for seed extensions in the estimated path of the seed and not everywhere. The seeds are then extended with compatible hits along the search roads using a combinatorial Kalman filter [70] to produce a set of track candidates. Following this, an ambiguity resolution step is performed on this set to remove low-quality candidates sharing too many hits with higher quality ones⁴. The purified candidates are then fitted with a χ^2 method to obtain an estimation of the track parameters.

Finally, for each candidate, a track extension in the TRT is attempted to improve the quality of the track. As previously, a search road is built in the TRT from the estimated entry point of the track candidate, following the estimated trajectory. The Kalman filter is used again to construct candidate extensions. The TRT hits in the extensions are then added to the corresponding track candidates. A refit is applied afterwards to the extended tracks to obtain the final track parameters estimation.

Secondary tracking chain, back-tracking pass In order to extend the track reconstruction acceptance to charged particles originating farther away from the beam line, a secondary tracking chain is run. Contrary to the first stage, it is a back-tracking pass using the hits not associated to any tracks reconstructed by the primary chain. It is used only in regions of interest (RoI) corresponding to energy deposits in the EM calorimeter. It starts by building sets of compatible TRT hits called segments that are in the vicinity of the RoI. Then, similarly as the primary chain, seeds are constructed from these segments with pairs of SCT hits close to the TRT and compatible with them. These seeds are extended to the rest of the SCT and Pixel sub-detectors using a search road and the Kalman filter. They are filtered by an ambiguity resolution step and the remaining ones are fitted to obtain new track candidates. Finally, the candidates are extended back into the TRT using the segments and fitted to obtain a new set of track parameters estimation.

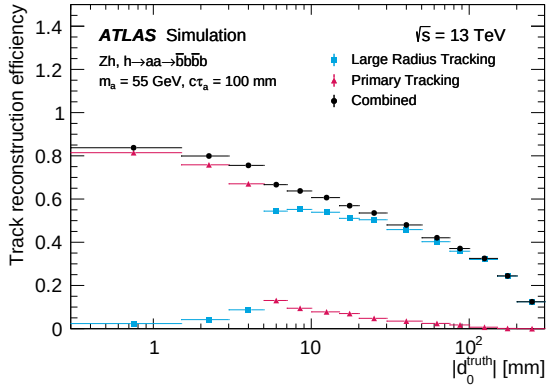
Large Radius Tracking A third tracking chain is applied to specifically reconstruct large d_0 tracks coming from the decay of long-lived particles. This last step uses the left-over hits from the primary and secondary chains. This tracking pass is mostly similar to the inside-out one but with wider search space in terms of impact parameters and loosened quality selections. Table 5.4 presents the most important selection and quality criteria for the primary tracking and LRT. It mainly shows the search space extension for the impact parameters allowing the LRT to reconstruct large radius tracks.

Performance The performance of the ATLAS track reconstruction software has been extensively tested in Ref. [68, 69]. The main performance of interest for the analysis presented in this thesis is the reconstruction efficiency evaluated on MC datasets. This quantity is defined by the fraction of charged truth particles to which a reconstructed track can be associated. Figure 5.5 shows this efficiency as a function of the $|d_0|$ of the truth particle, $|d_0^{\text{truth}}|$, and the radius of the production vertex of the truth particle, $R_{\text{prod}}^{\text{truth}}$, for the main (primary + secondary) tracking chain and the LRT. The efficiencies are measured on a MC dataset simulating a model with decaying BSM LLPs. It shows that the tracking efficiency is very good at low $|d_0^{\text{truth}}|$ and $R_{\text{prod}}^{\text{truth}}$, close to the beam line and the IP where the main tracking chain is optimised. It progressively decreases when going farther away from the beam line. It also shows the large gain in efficiency from the LRT alone above $|d_0^{\text{truth}}| = 5$ mm and above $R_{\text{prod}}^{\text{truth}} = 50$ mm, making its importance clear for LLP analyses.

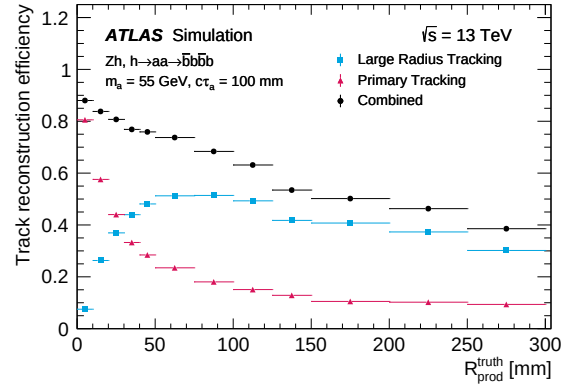
⁴ Track candidate quality is based on several criteria like the number of shared hits in the SCT and estimated impact parameters.

Figure 5.4: Summary table of the most important selection and quality criteria used for the primary and large radius tracking [69].

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



(a)



(b)

Figure 5.5: The primary + secondary (here shown as 'Primary'), LRT, and combined track reconstruction efficiencies in MC simulation for displaced charged particles produced by the decay of an exotic LLP. Efficiencies are shown as a function of the truth particle's (a) impact parameter $|d_0^{\text{truth}}|$ and (b) radius of production vertex $R_{\text{prod}}^{\text{truth}}$ [69].

5.3 Particle decay vertex reconstruction

From the set of reconstructed tracks, two reconstruction algorithms are used for the identification of the primary vertices, which are the interaction points in the pp collisions, and the secondary vertices, the late decay points of particles produced at the primary vertices. The former is crucial for pile-up mitigation techniques and the identification of the primary vertex at the origin of the event of interest. The latter allows reconstructing decay vertices of LLPs like b-hadrons or new exotic particles and is thus crucial for SM and BSM analyses.

5.3.1 Primary vertex reconstruction

The primary vertex reconstruction is a challenging task due to the high level of pile-up reached at the end of Run 2 and during Run 3, with up to 60 pp collisions per bunch crossing (as outlined in Fig. 4.12). The reconstruction algorithm used by ATLAS is called the *adaptive multi vertex fitter* (AMFV) and is detailed in Ref. [71] along with its performance. Only a simplified summary will be given here.

The AMFV iteratively builds vertex candidates from a quality-based selected set of reconstructed tracks. From a seeding pool containing tracks not associated to any vertex candidates, the most likely position of a new vertex, a seed, is estimated from the track density along the beam line. The track density is measured using the impact parameters of the tracks in the seeding pool, a region of over-density corresponding most likely to the position of a pp interaction vertex. After a seed is found, neighbouring tracks are considered for assignment, creating a vertex candidate, based on their consistency of being products of the same vertex. At this step, the complete set of tracks is considered (i.e. not only the ones unassigned to any candidates), meaning that tracks can be associated to multiple vertex candidates with the use of varying weights. These weights measure the compatibility of a track with each vertex candidate. Then the parameters and weights of the assigned tracks are used to fit the position of the vertex candidate. Other vertex candidates linked to the newly considered candidate (through shared tracks) are also simultaneously fitted, and all track weights are recalculated. Afterwards, the new vertex candidate can either be accepted and added to the set of candidates for the next iterations or rejected. The criteria to be accepted are:

- to have at least two compatible tracks from the seeding pool⁵,
- to have a fitted position outside 3σ of any other vertex candidates, based on the fit errors of either candidates.

If the vertex candidate is accepted, all tracks compatible with it are removed from the seeding pool and the candidate will be used in the next iterations. If it is rejected, only the most compatible track is removed. The algorithm stops when there are less than two tracks in the seeding pool or it is not able any more to find a seed. At the end of the AMFV, all vertex candidates are taken as reconstructed primary vertices.

The algorithm outputs a collection of fitted vertices position to which compatible tracks are linked. Usually, the hard-scatter (HS) vertex at the origin of the event of interest is selected as the reconstructed primary vertex with the greatest squared sum of the transverse momentum of the associated tracks:

$$\sum_{\text{trk} \in \text{PV}} p_{\text{Trk}}^2$$

⁵ It can not be only formed of tracks already compatible with previous vertex candidates.

5.3.2 Secondary vertex reconstruction

The second vertex reconstruction algorithm was developed to identify and reconstruct secondary displaced decay vertices of LLPs inside the inner detector and outside the region of interest of the AMFV algorithm. Some SM particles (like b-hadrons) or BSM particles can be produced promptly at the IP and travel some distance in the inner detector before decaying to new SM particles. If some of the produced particles are charged and their tracks are reconstructed, it is then possible to identify the position of the decay by looking for intersection points between tracks. This reconstruction algorithm is also of interest in the frame of BSM searches for new exotic LLPs with non-zero lifetime leading to decays in the inner tracker.

The secondary vertex reconstruction algorithm [72] is based on a set of preselected reconstructed tracks. From this set, two-track seeds are identified based on the assessed compatibility of all possible pairs of tracks. The seeds are then merged to form multi-track vertices and can be combined with nearby vertices if possible. Finally, tracks not retained for vertex seeding are attached to compatible vertices.

Track selection for seed finding As stated, tracks entering the identification of seed vertices are preselected from the total pool of reconstructed tracks (standard tracking + LRT). This is to reduce the combinatoric by removing low quality tracks and thus improve the reconstruction efficiency. The main selections are on the transverse momentum, requiring a $p_T > 1$ GeV (i.e. the lower cut for LRT tracks) and no association to any primary vertices. More specific quality-based selections are applied based on the number of hits in the different sub-detectors.

Two-track seed finding From the pool of preselected tracks, all possible pairs of tracks with $d_0 > 2$ mm are formed. Each pair is then used to compute an approximated vertex position based on the track parameters measured at the perigee. Pairs of tracks with small impact parameters measured at the estimated vertex position⁶ are used for a vertex candidate position fitting algorithm [73]. If the fit can not find a viable solution the vertex candidate is rejected. If the fit is successful, both tracks constitutive of the vertex candidate are then required to fulfil the hit-pattern consistency. It requires that the track has hits in tracker layers with radii greater than the estimated vertex position radius and no hits in precedent layers. Here, tracks originating from secondary vertices are supposed to propagate from the inside to the outside of the detector. This requirement is applied, among other reasons, to avoid the reconstruction of vertices from uncorrelated crossing tracks. Vertex candidates with both tracks fulfilling this requisite are retained as vertex seeds.

Multi-track vertex forming During the seed finding stage, no requirement on track exclusivity is applied. Therefore, tracks originating from the same LLP decay can form multiple vertex seeds or a single track can be used in multiple vertex seeds. To better reconstruct the LLP decay vertices, multiple vertex seeds are merged using the track compatibilities between the seeds. An undirected graph network is constructed where nodes represent tracks and segments represent the vertex seeds. This graph is called a compatibility graph, nodes (tracks) connected through edges (vertex seeds) are compatible to form a multi-track vertex. All group of fully compatible nodes are extracted and each of these groups of tracks is used to define multi-track vertices. Fits are then applied to estimate the position of the multi-track vertices.

⁶ The track impact parameters are here re-calculated by changing the reference point to be the estimated vertex position.

Vertex merging After the multi-track vertex forming, multiple reconstructed vertices can still originate from the same LLP decay. As for the previous step, vertices are again allowed to be merged into a single vertex. Two vertices are considered compatible for merging if the difference in position between the two is within 10σ (where σ is defined as the uncertainty on the difference in position). For all group of compatible vertices, the merging is done in ascending order of associated track multiplicity, starting from the lowest multiplicities and trying to merge them with the highest multiplicities first. The merging compatibility is further assessed with different tests and all pairs of vertices fulfilling one of these tests are merged. The tracks from both vertices are then used to fit the new vertex position. Finally, any remaining vertices within 1 mm from another are forced to merge.

Track attachment At last, all the unselected tracks for the seed finding step are used to be associated to compatible reconstructed vertices. A track is associated to a vertex if its impact parameters recalculated at the vertex position are compatible with the vertex (i.e. close to zero impact parameters). Here again, multiple associations are possible and are avoided by limiting the tracks to one vertex association, the highest track multiplicity one. This track attachment is designed to accept as much track as possible by loosening the requirements previously applied to the track selected for seeding. These associations aim at improving the vertex kinematics measurements and so the characterisation of the LLP decay. After all compatible tracks are attached, for each vertex a final fit is conducted with all the associated tracks and all the associated track parameters are re-calculated with respect to the position of the vertex.

This algorithm, outputs a collection of fitted reconstructed secondary vertex positions along links to their associated tracks and kinematic properties (number of tracks, invariant mass, total p_T).

Performance The performance of the secondary vertex reconstruction algorithm is presented in details in Ref. [72]. Only the total reconstruction efficiency as a function of the truth LLP decay radius for different LLP models is presented in Fig. 5.6. This plot shows the range in decay radius for which the reconstruction is performant and the dependency on the mass of the decaying LLP. The three models presents different processes of production of LLPs (using either supersymmetry, the exotic decay of the Higgs boson or a heavy neutral lepton) but, in the context of secondary vertex reconstruction, they differ mainly in the mass of the decaying LLP, ranging from tens of GeV up to the TeV scale. The resulting efficiencies show a large dependency on this mass explainable by the number and energy of tracks produced from the LLP decay being highly correlated to the mass of the LLP. The secondary vertex reconstruction shows better performance for vertices with large track multiplicities with $p_T > 1$ GeV.

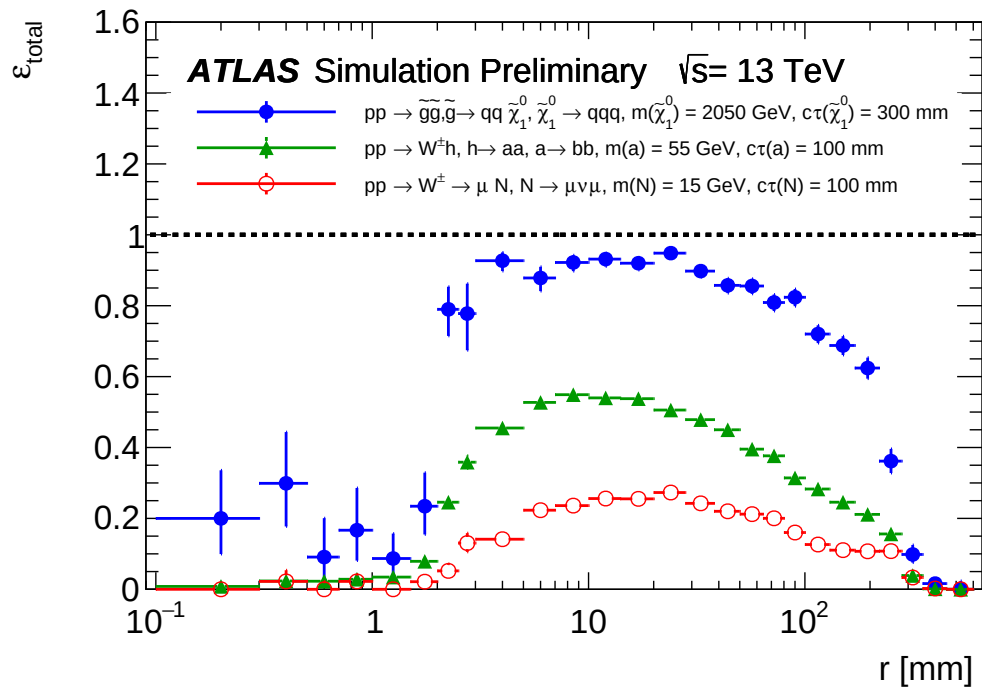


Figure 5.6: Secondary vertex reconstruction efficiency measured in MC simulation as a function of the truth LLP decay radius for three LLP models [72].

5.4 Jet reconstruction

Jet refers to the spray of particles that results from the emission of an energetic quark or gluon. They are abundantly produced at the LHC during pp collisions and are of crucial importance for many analyses either because they compose the main signal signature or constitute a source of background. It is therefore necessary to precisely and consistently reconstruct these physics objects in order to measure their angular positions in the detector, their energies and their masses.

5.4.1 Jet clustering

An energetic quark or gluon emitted during a pp collision cannot propagate freely because of the colour confinement phenomena of the strong force. As described in Section 2.2, it first initiates new emissions of gluons and pair-produced quarks during a parton shower, thereby losing energy until it reaches a certain energy scale (Λ_{QCD}). Then the emitted quarks and gluons confine to form colourless hadrons during the hadronisation. The produced spray of hadrons usually forms a cone-shaped structure centred around the direction of the initiating quark or gluon. In an ideal case where only one quark or gluon is emitted and thus one jet is produced, reconstructing this jet, its angular position, energy and mass is easy. The jet kinematics would be obtained from the four-momentum sum of all the produced particles during hadronisation. However, at the LHC, many more particles are usually produced during pp collisions, in the main HS collision or the underlying collisions, leading to many more emissions and physics objects overlapping each others. Moreover, due to pileup, multiple pp collisions can occur in the same event leading to an even busier environment. It is thus crucial to be able to reconstruct a jet in such a busy environment by only grouping particles belonging to the jet.

This is the purpose of clustering algorithms which, from a set of particles four-momenta, define groups of particles that most likely originate from the same jets. These algorithms are characterised by their choices of clustering rules and corresponding parameters (mainly the angular extent of the jet) and thus the obtained reconstructed jets depends on these choices. However, whatever choice of clustering method, these algorithms need to be collinear and infrared safe⁷, such that they can be implemented in theoretical QCD calculations. This specification is crucial for physics analyses to confront jet production cross-section measurements to theoretical predictions. They are also largely sensible to pileup and underlying event particles emitted in the same region. Therefore, the clustering algorithms generally need to be complemented by pileup mitigation and grooming techniques either at the input-particle or jet level.

In ATLAS, the jet clustering is done using the FASTJET software [74] and the anti- k_t sequential algorithm [75]. The anti- k_t algorithm regroups particles based on two criteria, the angular distance between particles ΔR (defined in Eq. 4.1) and their individual transverse momentum. It takes as inputs a set of four-vectors which can correspond to either truth particles from MC simulations, energy deposits in the calorimeter, charged particle tracks or a combination of the last two. This will be detailed later and the term particle will generically be used in the following discussion. The algorithm starts by calculating, from a set of particle four-vectors, the following 'distances':

- for each pair of particles i and j :

$$d_{ij} = \min\left(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}\right) \frac{\Delta R_{ij}}{R}$$

⁷ Random collinear or soft emissions during the parton shower must not alter the jet clustering.

- for each particle i :

$$d_{iB} = \frac{1}{p_{Ti}^2}$$

where R is a free parameter that defines the angular opening of the final jet. The algorithm then searches for the minimum distance in the list of (d_{ij}, d_{iB}) . If the minimum is one of the d_{ij} distances then the particles i and j are combined into a new particle by adding their four-vectors. The newly combined particle is added to the set of particles, and i and j are removed. On the other hand, if the smallest distance is one of the d_{iB} , then the particle i is considered as a final jet (it can be a combination of multiple particles) and is removed from the set of particles. The process is repeated until no more particles are left. This clustering algorithm regroups particles around the most energetic ones (due to the $1/p_T^2$ term) and produces mostly symmetrical cone-shaped jets with an angular opening R . It is by definition collinear and infrared safe. All jet related variables are then computed from the set of four-vectors that compose them. The quantities (E, m, ϕ, η, p_T) are obtained by summing the four-vectors and represent the jet kinematics. The jet observables describing the jet internal structure (generally referred to as the jet substructure) are computed from the constituents four-vectors themselves, capitalising on their angular and energy distribution. These substructure observables are usually used for the tagging of the particles initiating the jets (e.g. light quarks, gluons, b-quarks, top quarks, ...).

Other clustering algorithms exist which mainly vary the exponent of the p_T terms in d_{ij} and d_{iB} to change the order of the clustering between hard and soft emission, thus changing the jet angular shapes. Two commonly used variants of the anti- k_t algorithm are:

- the k_t algorithm [76] where the clustering is driven by $\min(p_{Ti}^2, p_{Tj}^2)$, thus grouping soft particles first,
- the *Cambridge-Aachen* (C/A) [77] algorithm where the clustering is driven only by the angular separation (the p_T exponent is set to 0) thus grouping adjacent particles first.

These two algorithms are commonly used for specific calculations of observables describing the jets substructure.

5.4.2 Jet angular opening: small- R and large- R jets

In ATLAS, the main jet collections used for physics analyses are reconstructed with either $R = 0.4$ or $R = 1.0$. The former is used to reconstruct standard QCD jets produced by the emission of a light quark (u, d, s, c, b) or gluon and are called small- R jets. The latter, called large- R jets, are mainly used to study the hadronic decays of massive boosted particles like W/Z and Higgs bosons and top quarks, which produces jets with larger spread of particles and thus larger angular opening.

5.4.3 Jet input constituents

As stated before, the jet clustering algorithm takes as input a list of four-vectors. These are the four-vectors of the jet input *constituents*. The choice of constituents determines the type of jets and influences largely the performance of jet reconstruction in terms of angular and energy measurements, corresponding resolutions and pileup mitigation. While most analyses use the recommended most performant jet constituents, some

more exotic analyses have to make a more thoughtful choice depending on the explored signatures, this will be discussed later in Section 7.2.

5.4.3.1 Pure calorimeter constituents: topological cell clusters

Jet constituents formed only from calorimeter measurements are called *Topological cell clusters* [78] and commonly shortened to *topo-clusters*, they are the basic blocks for energy measurements in ATLAS. Topo-clusters are aggregations of topologically-connected calorimeter cell signals and aim at extracting significant signals from particle energy depositions over electronic noises and sources of fluctuations like pileup. The high granularity of the ATLAS calorimeters allows reconstructing three-dimensional energy clusters produced by particle showers. However, with the high pileup rate and dense environment of the LHC, a topo-cluster does not always contain only one particle energy deposition. A particle energy can be split in several clusters or a cluster can contain contributions from different particles.

Cluster formation is based on a three-dimensional spatial signal-significance pattern reconstruction. It is based on each cell significance $\varsigma_{\text{cell}}^{\text{EM}}$ defined as the ratio between the cell measured signal $E_{\text{cell}}^{\text{EM}}$ and the cell expected average noise $\sigma_{\text{noise,cell}}^{\text{EM}}$:

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

$\sigma_{\text{noise,cell}}^{\text{EM}}$ is the quadratic sum of the electronic noise and noise created by pileup. The pileup noise is created by the contributions to the calorimeter response of in-time and out-of-time pileup, it is estimated each running year from the expected average number of interaction per bunch crossing $\langle \mu \rangle$. Both cell signals and significances are measured at the electromagnetic (EM) scale without correction for the non-compensating property of the ATLAS calorimeters for hadronic showers. The cell's electric signal to energy conversion is adapted to the energy deposition of electron and photons and are thus well measured from the EM showers in the EMCal. However, hadrons measured in the HCal are subject to energy losses during hadronic showers as some emitted low energy hadrons can be absorbed by the high-density absorbers before being sampled. These energy losses for the hadrons are not compensated for during the signal to energy conversion of the cells. This scale correspond in fact to the raw measurements from the calorimeters.

Cells are clustered together based on three equations and three parameters $\{S, N, P\}$:

$$|E_{\text{cell}}^{\text{EM}}| > S\sigma_{\text{noise,cell}}^{\text{EM}} \Rightarrow |\varsigma_{\text{cell}}^{\text{EM}}| > S \quad (5.1)$$

$$|E_{\text{cell}}^{\text{EM}}| > N\sigma_{\text{noise,cell}}^{\text{EM}} \Rightarrow |\varsigma_{\text{cell}}^{\text{EM}}| > N \quad (5.2)$$

$$|E_{\text{cell}}^{\text{EM}}| > P\sigma_{\text{noise,cell}}^{\text{EM}} \Rightarrow |\varsigma_{\text{cell}}^{\text{EM}}| > P \quad (5.3)$$

They define three categories of signal significance for cells in decreasing order with $\{S, N, P\}$ taken per default as $\{4, 2, 0\}$.

The first step of cluster formation is the selection of all cells (i.e. seeds) passing Eq. 5.1 to form proto-clusters. Then for each seed in decreasing $\varsigma_{\text{cell}}^{\text{EM}}$ order: all directly adjacent cells regarded as neighbouring cells (in the same calorimeter layer or in adjacent layers), passing Eq. 5.2 and Eq. 5.3 are collected and associated to the proto-cluster. For neighbouring cells passing Eq. 5.2, the same method is applied,

adjacent cells passing Eq. 5.2 and Eq. 5.3 are collected. This procedure is iteratively repeated until the last set of collected neighbours contains cells passing only Eq. 5.3. At this point the clustering for this proto-cluster is stopped and the clustering passes to the next seed. During this procedure two special cases can occur:

- if a neighbouring cell is a seed (i.e. passes Eq. 5.1) then the two proto-clusters are merged,
- if a neighbouring cell is already attached to another proto-cluster and if it passes Eq. 5.2 then the two proto-clusters are merged, else it is shared between the two seeds.

This procedure allows to only consider clusters with at least one highly significant signal cell and so removes low (soft) isolated energy depositions. The output is a collection of topo-clusters measured at the EM scale called EMTopo clusters which is available as jet input constituent.

Negative signal cells and pileup estimation The equations selecting significant cells use absolute values meaning they can select cells with negative signals. Negative signal cells are possible for the LAr calorimeter (EMCal) as its response pulse shape is composed of a negative part as shown in Fig. 5.7 so that its integral over time values to zero. This was chosen so that signal residuals (soft emissions) in a calorimeter cell from in-time and out-of-time pileup would be cancelled on average. Negative signal cells are mainly the product of energetic pileup contributions but also electronic noises.

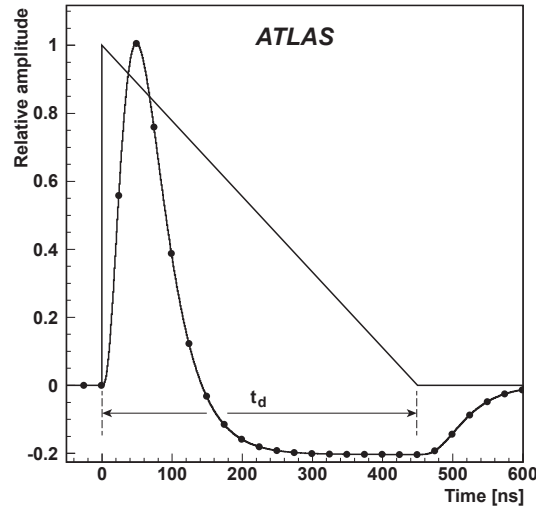


Figure 5.7: The pulse shape in the ATLAS LAr calorimeter. t_d is the characteristic time of the pulse called drift time, $t_d \approx 450$ ns. The points are separated by 25 ns, corresponding to the bunch crossings [78].

The use of absolute values in the clustering algorithms thus means that highly significant negative signal cells can be seeds to clusters and negative signals can contribute in the clustering and be associated to globally positive clusters.

Clusters seeded by negative signal cells will most likely also have a negative total energy. The number of this type of clusters can be used as an estimator for the rate of out-of-time pileup. Generally, they are not used for physics object reconstruction such as jet, as they are considered to not contribute to the actual signal generated by the event of interest. On the other hand, out-of-time pileup contribution can

also produce positive clusters from particles generated by closer in time bunch crossings (from Fig. 5.7, within 100 ns). Negative clusters can hence also be used in global event observables to cancel the positive contributions of out-of-time pileup. For example, some methods of evaluating the missing transverse momentum using only the calorimeter information may include these negative signals.

The clustering of negative signal cells within positive clusters is also included as it provides a noise suppression capability to topo-clusters by cancelling random negative and positive noise fluctuations.

Local hadronic calibration and signal corrections The resulting topo-clusters can be directly used as constituents for jet reconstruction. However, all signal cells are taken at the EM scale without any particular treatment meaning that the reconstructed topo-clusters are also at the EM scale. This is problematic as the energy measurement of hadrons is different from the measurement of photons and electrons leading to non-linearities in the calorimeter cell responses. Jets with different EM and HAD composition would thus have different detector response behaviours. Moreover, both ECal and HCal have inherent measurement inefficiencies due to energy losses in inactive/dead materials. To correct for such effects, ATLAS developed a cell-per-cell calibration to the topo-clusters called *Local Cell Weighting* (LCW). This calibration is derived from a study of single-pion calorimeter response in MC simulation. The LCW calibration is based on the differentiation of topo-clusters initiated by EM showers or HAD showers as the cells response depends widely on the nature of the energy deposit. This classification procedure uses the information available on the cluster like its position in the calorimeter (ϕ, η) and its longitudinal and transverse shapes, and outputs a number interpreted as the probability of it being generated by an EM shower $\mathcal{P}_{\text{clus}}^{\text{EM}}$. Then, two different cell calibration factors for EM and HAD showers are derived from the MC simulation. The global cell calibration is taken as the linear combination of the two using $\mathcal{P}_{\text{clus}}^{\text{EM}}$ and $1 - \mathcal{P}_{\text{clus}}^{\text{EM}}$ as coefficients. Other corrections are also applied to compensate for energy losses from out-of-cluster (missed) cells and dead material. All the calibration factors are finally combined to obtain the total cell calibration factor $w_{\text{cell}}^{\text{calib}}$. The calibrated cluster energy is then computed as:

$$E_{\text{clus}}^{\text{calib}} = \sum_{i \in \text{cluster}} w_{\text{cell},i}^{\text{calib}} * E_{\text{cell},i}^{\text{EM}} \equiv E_{\text{clus}}^{\text{LCW}}$$

After this calibration, the cluster is said to be at LCW scale. This second collection of calibrated topo-clusters called LCTopo clusters can also be used for jet reconstruction.

5.4.3.2 Track-calorimeter combined constituents

Early ATLAS analyses used reconstructed jets build only from calorimeter deposits through the topo-clusters. Although the topo-clusters shows excellent energy resolution that increases with the energy, they lack precision in the position of deposits to resolve angular separations due to the finite segmentation of the cells. Tracks however, have an opposite behaviour: they present a high precision in particle trajectory measurement combined with a precise transverse momentum measurement at low energy (up to $p_T \simeq 100$ GeV). Both objects have complementary performance. Therefore, combining topo-clusters and tracks offers the possibility to improve the energy resolution for low-energy particles and the accuracy on the position of energy deposits, thus improving the global reconstruction performance of jets. On top of that, the addition of tracks and their potential link to primary vertices can allow to remove constituents originating from pileup vertices, mitigating in-time pileup emissions. Based on this observation, the ATLAS Collaboration developed several new jet constituents collections combining topo-clusters and tracks in different ways.

Particle Flow Objects The usage of both tracks and topo-clusters for jet reconstruction cannot be done directly as particles that produce tracks most likely also deposit energy in the calorimeters. So using both the track and topo-cluster collections would lead to double counting of particle energies. A first strategy to avoid this is to use a matching between tracks and topo-clusters and subtract the energy of the tracks from the matched topo-clusters. This strategy is implemented in the *Particle Flow* (PFlow) algorithm [79, 80]. The PFlow algorithm workflow is presented in Fig. 5.8. It uses a subset of selected tracks and the collection of topo-clusters at the EM scale.

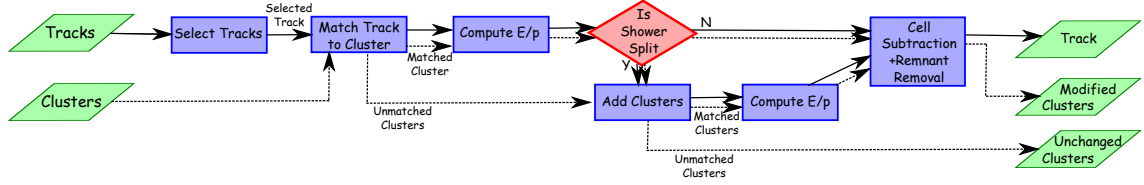


Figure 5.8: The Particle Flow algorithm [79].

The first step of the algorithm is the track selection. A tight selection is applied to keep only clean reconstructed tracks with at least 9 hits in the Pixel and SCT detectors and no missing hits in the Pixel detector. Additionally, tracks are required to have $|\eta| < 2.5$ and a transverse momentum $p_T > 0.5$ GeV. These selections allow to lower the computational complexity of the track-to-topo-cluster matching while keeping tracks that would not create topo-clusters (because of the energy thresholds needed to seed a cluster) or would not even reach the calorimeter (because of a small curvature radius). Furthermore, high-energy tracks ($p_T > 100$ GeV) are also removed for two reasons. They are generally produced in dense environment with straighter trajectories and small separations with other components of the hadronic showers making them poorly isolated. This poor isolation would lead to decreased performance for the next steps of the algorithm. Moreover, at these energies, the track energy resolution become poorer than the calorimeter energy resolution.

Each track is then considered in a matching attempt to a single topo-cluster. The matching is based on the angular distance $\Delta R'$ between a topo-cluster barycentre and the track position extrapolated to the second layer of the EM calorimeter. $\Delta R'$ is given by:

$$\Delta R' = \sqrt{\left(\frac{\Delta\phi}{\sigma_\phi}\right)^2 + \left(\frac{\Delta\eta}{\sigma_\eta}\right)^2} \quad (5.4)$$

and takes into account the angular topo-cluster widths (σ_ϕ, σ_η) taken as the standard deviation in ϕ and η of the displacements of the topo-cluster's cells regarding its barycentre. For example:

$$\sigma_\phi = \sqrt{\sum_{i \in \text{clus}} (\phi_i^{\text{cell}} - \phi^{\text{barycentre}})^2 - \left(\sum_{i \in \text{clus}} \phi_i^{\text{cell}} - \phi^{\text{barycentre}}\right)^2}$$

For each track, the topo-cluster with the closest $\Delta R'$ is considered as the matched topo-cluster. If there are no clusters within $\Delta R' = 1.64$ then the track is not matched and the charged particles at the origin of the track is considered not to have deposited energy in the calorimeters, it can then directly be used as a jet constituent.

After that, for each pair of (track, topo-cluster) it is necessary to determine the energy that was deposited by the particle corresponding to the track in the topo-cluster $\langle E_{\text{dep}} \rangle$. From $\langle E_{\text{dep}} \rangle$ it will be then possible to subtract the right energy to the matched topo-clusters. $\langle E_{\text{dep}} \rangle$ is computed from:

$$\langle E_{\text{dep}} \rangle = p^{\text{trk}} \left\langle \frac{E_{\text{ref}}^{\text{clus}}}{p_{\text{ref}}^{\text{trk}}} \right\rangle \quad (5.5)$$

It is a measure of the expected deposited energy from a particle with momentum p^{trk} on average. $\langle \frac{E_{\text{ref}}^{\text{clus}}}{p_{\text{ref}}^{\text{trk}}} \rangle$ is the average ratio between the energy deposited in the cluster and the momentum of the depositor track, it is in fact the mean response of a cluster to a track. It is measured from single-pion simulated samples, where $E_{\text{ref}}^{\text{clus}}$ (the energy of clusters produced by the pion) is taken as the sum of the topo-clusters in a cone of $\Delta R = 0.4$ around the pion track position. It was determined that this cone size is large enough to contain the showers of most particles. Furthermore, $\langle \frac{E_{\text{ref}}^{\text{clus}}}{p_{\text{ref}}^{\text{trk}}} \rangle$ is measured in bins of track η and p_T and the layer of highest energy density of the topo-cluster, in order to take into account the detector geometry and shower development variations.

Until now, it was implicitly assessed that a particle and its associated track only deposit energy in its single matched topo-cluster. However, as discussed in the previous subsection, it is possible for a particle to deposit energy in several topo-clusters. In such cases, it is then necessary to determine all the topo-clusters that need to be associated to the same (track, topo-cluster) pair, in order to subtract the track energy $\langle E_{\text{dep}} \rangle$ to the complete system of clusters. The discrimination between single and multiple topo-clusters deposits is made using the significance between the energy of the topo-cluster originally matched to the track E^{clus} and $\langle E_{\text{dep}} \rangle$:

$$S(E^{\text{clus}}) = \frac{E^{\text{clus}} - \langle E_{\text{dep}} \rangle}{\sigma(E_{\text{dep}})} \quad (5.6)$$

where $\sigma(E_{\text{dep}})$ is the standard deviation of the $E_{\text{ref}}^{\text{clus}}/p_{\text{ref}}^{\text{trk}}$ ratio. If $S(E^{\text{clus}}) < -1$ the energy deposit is considered to be split and so all topo-clusters within $\Delta R = 0.2$ around the track position are gathered in the same (track, topo-clusters) system. After this step, the formed (track, topo-clusters) systems contain one track and one or multiple topo-clusters.

Subsequently, with the estimated energy depositions $\langle E_{\text{dep}} \rangle$ for each system of track and matched topo-clusters, the subtraction of the track energies to the topo-clusters is applied. If $\langle E_{\text{dep}} \rangle$ is greater than the energy sum of the topo-clusters in the system then all the clusters are removed. If it is not the case, a cell-by-cell subtraction is performed, removing iteratively cells contained in the set of matched topo-clusters until the removed energy reaches $\langle E_{\text{dep}} \rangle$. However, this procedure is not always applied depending on the track transverse momentum and the surrounding environment. In high- p_T jets, energetic particles are often grouped in small areas at the core of the jets and are poorly isolated. In these areas, it is thus difficult to accurately subtract the topo-clusters energies associated to a track. Therefore, a selection is made to only apply the subtraction when the track does not fall in areas with significant energy depositions from other particles. If the summed calorimeter cell's energies $E_{\Delta R=0.15}^{\text{clus}}$ in a cone of size $\Delta R = 0.15$ around the position of the track satisfies

$$\frac{E_{\Delta R=0.15}^{\text{clus}} - \langle E_{\text{dep}} \rangle}{\sigma(E_{\text{dep}})} > 33.2 \times \log_{10}(40\text{GeV}/p_T^{\text{trk}}) \quad (5.7)$$

then the subtraction is not applied. In these cases, the track is still used for its angular measurement but not its energy. This allows for a better transition between the use of track and topo-cluster energy measurements.

Finally, for each (track, topo-clusters) system, if the remaining cell energies after subtraction is consistent with shower fluctuations, meaning consistent with the width of $E_{\text{ref}}^{\text{clus}}/p_{\text{ref}}^{\text{trk}}$, it is assumed that the set of matched topo-clusters results from a single particle deposition. The remaining cells are thus considered to be only the product of shower fluctuations and therefore removed. In the case where the remaining cells are not consistent with shower fluctuations, it is on the contrary assumed that the set of matched topo-clusters was produced by multiple close-by particles. The remaining topo-clusters are then retained.

At the end of the PFlow algorithm a collection of PFlow objects (PFOs) is obtained containing:

- all selected unmatched tracks and matched tracks with or without subtraction (both are referred as *charged PFOs*),
- all unmatched topo-clusters and the remaining ones that survive the subtraction step (referred as *neutral PFOs*).

This collection of PFOs represents the reconstructed events without double counting between tracks and topo-clusters. The PFOs can then be used as input constituents for jet reconstruction, yielding better angular and energy resolution at low energy than EMTopo or LCTopo constituents but also pileup mitigation opportunities, both through the use of tracks.

Track-CaloClusters The PFlow constituents show the best performance for reconstructing jets at low energy. This makes these constituents good candidates to reconstruct low-energetic hadronic processes like standard QCD jets produced from light-quarks or gluons. However, highly-energetic hadronic processes produced for example by boosted hadronic decays (W/Z, Higgs or top decays) are also of great interest for ATLAS. These kinds of processes are challenging to reconstruct due to the boosted topology resulting in dense environments where particles are well collimated. The PFlow constituents in these cases are not optimal due to the treatments of tracks in dense environments. Therefore, the ATLAS Collaboration developed a second algorithm for combining tracking and calorimeter information called *Track-CaloClusters* (TCC) [81]. This algorithm aims at maximally exploiting the strengths of the tracks and topo-clusters at higher energy than the PFlow algorithm. Tracks at higher energy show degraded p_T resolution but, as they get straighter, show increased angular resolution when extrapolated to the calorimeter. Hence, TCC aims at reconstructing high- p_T constituents benefiting from both the excellent calorimeter energy resolution and the tracker spatial resolution by building 4-vectors where:

- the energy scale (E, m, p_T) is measured from the topo-clusters,
- the angular coordinates (ϕ, η) are measured from the track parameters.

Similarly as PFlow, the TCC algorithm is based on the matching between tracks and topo-clusters. Though, instead of trying to subtract energy of tracks from the topo-clusters, here the tracks are only used to improve the angular measurement of the energy deposits without the use of track momentum information.

The TCC algorithm uses all topo-clusters at the LCW scale and all tracks passing loose quality criteria, from both the HS primary vertex and pileup vertices. It starts by attempting to match all tracks to every topo-clusters using the track parameters extrapolated to the calorimeter and the topo-clusters barycentre angular positions. Only tracks with extrapolated parameters uncertainty smaller than any cluster widths ($\sigma^{\text{clus}} = \sqrt{\sigma_{\phi}^{\text{clus}^2} + \sigma_{\eta}^{\text{clus}^2}}$) are considered for matching. A track-cluster pair is defined whenever their angular separation is lower than the sum of their angular uncertainties:

$$\Delta R < \sqrt{\sigma^{\text{clus}^2} + \sigma^{\text{trk}^2}}$$

where σ^{trk} is the uncertainty on the track parameters extrapolated to the calorimeter. Track and topo-cluster pairs are used as seeds for the TCC building. If the track is originating from a pileup vertex, then the pair is not used for seeding. At this point, multiple tracks can be matched to the same topo-cluster and/or multiple topo-clusters can be matched to the same track. Therefore, the main task of the TCC algorithm is to share the energy from multiple topo-clusters into several TCC constituents, each one with a single associated track.

The following steps are performed to build the TCC-4-vectors from all the matching pairs. When necessary, the calculation needs to take into account possible sharing problems between tracks and topo-clusters. Fig. 5.9 shows the different possibilities resulting from the matching between tracks and topo-clusters and the corresponding reconstructed TCC objects. Each number ① represents a TCC object, red lines are tracks t_i , and yellow (EMCal) and blue (HCal) boxes are topo-clusters c_i .

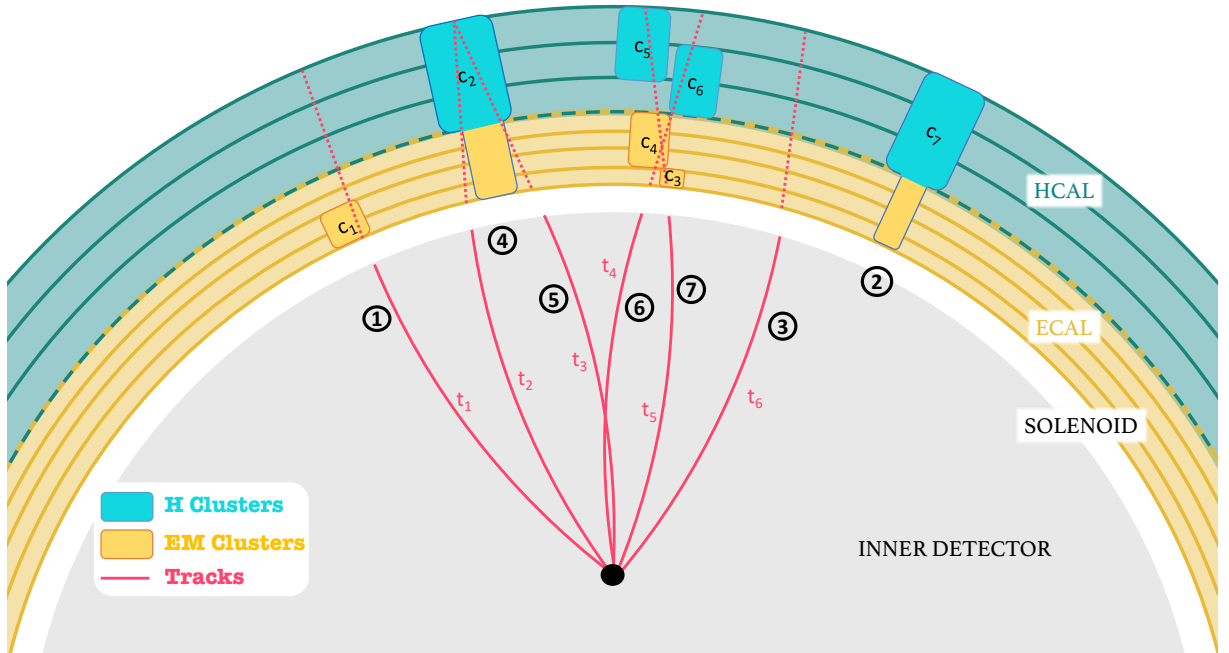


Figure 5.9: Schematic showing different possibilities of track to topo-clusters matching and the corresponding TCC object building [81].

The first three possibilities are the easiest as no special treatments are needed. TCC ① constitute the basic TCC object where one track is matched to one topo-cluster. In this case, the track t_1 is used for the

angular coordinates of the TCC 4-vector (η^{t_1}, ϕ^{t_1}) and the topo-cluster c_1 is used for the TCC energy scale ($p_T^{c_1}, m^{c_1}$). TCC ③ 4-vector is then constructed as:

$$\text{TCC}_{\textcircled{3}} = (p_T^{c_1}, \eta^{t_1}, \phi^{t_1}, m^{c_1} = 0) \quad (5.8)$$

Here the mass of the TCC is set to 0 as the mass of a single topo-cluster is defined to be zero in ATLAS. The TCCs ② and ③ are cases where, respectively, a topo-cluster has no matched track and a track has no matched topo-cluster. In both cases, the TCC 4-vector is then build from its single constitutive isolated object 4-vector, a topo-cluster or a track. Eq. 5.9 and 5.10 give the corresponding 4-vectors:

$$\text{TCC}_{\textcircled{2}} = (p_T^{c_7}, \eta^{c_7}, \phi^{c_7}, m^{c_7} = 0) \quad (5.9)$$

$$\text{TCC}_{\textcircled{3}} = (p_T^{t_6}, \eta^{t_6}, \phi^{t_6}, m^{t_6} = 0) \quad (5.10)$$

More complex possibilities arise when either multiple tracks are matched to the same topo-cluster, multiple topo-clusters are matched to the same track or multiple tracks are matched to multiple topo-clusters (④ and ⑤, ⑥, and ⑦, respectively). All these situations need to take into account sharing of the cluster energies. The TCC building procedure, as mentioned, builds one distinct TCC object for each seed track originating from the HS primary vertex. For each object the track coordinates are used but the energy scale needs to be adapted by sharing the topo-cluster energies between the different matched tracks. First, for each track seeding a TCC all matched topo-clusters are gathered. Then the energy sharing follows three rules:

- each topo-cluster c matched to the same seed track needs to contribute to the TCC object proportionally to its p_T fraction over all matched topo-clusters,
- each topo-cluster c matched to multiple track seeds t needs to contribute to each TCC objects proportionally to the p_T fraction of each track t over all matched tracks,
- each of those track t contribution to cluster c needs to be weighted by the p_T fraction of the topo-cluster c over all clusters matched to the track t .

The energy splitting is thus always based on p_T fractions between either multiple topo-clusters or multiple tracks. This procedure allows to never directly compare the energy measurements from the tracker and the calorimeter as only scale factors from one detector is used to scale the other. It also does not need a precise measurement of the track p_T as only ratios are used. For a simple example, this procedure applied to TCCs ④ and ⑤, where two seeding tracks are matched to the same topo-cluster, would give:

- the topo-cluster c_2 energy is split between track t_2 and t_3 following the p_T fraction of both tracks,
- the TCC seeded from t_2 takes the following p_T from c_2 :

$$p_T^{c_2} \frac{p_T^{t_2}}{p_T[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]}$$

- the TCC seeded from t_3 takes the following p_T from c_2 :

$$p_T^{c_2} \frac{p_T^{t_3}}{p_T[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]}$$

where $p_T[\mathbf{p}^a + \mathbf{p}^b]$ corresponds to the p_T of the 4-vector sum of particle momenta a and b . The corresponding TCC 4-vectors are thus given by 5.11 and 5.12:

$$\text{TCC}_{\textcircled{4}} = \left(p_T^{c_2} \frac{p_T^{t_2}}{p_T[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]}, \eta^{t_2}, \phi^{t_2}, m^{c_2} \frac{p_T^{t_2}}{p_T[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]} = 0 \right) \quad (5.11)$$

$$\text{TCC}_{\textcircled{5}} = \left(p_T^{c_2} \frac{p_T^{t_3}}{p_T[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]}, \eta^{t_3}, \phi^{t_3}, m^{c_2} \frac{p_T^{t_3}}{p_T[\mathbf{p}^{t_2} + \mathbf{p}^{t_3}]} = 0 \right) \quad (5.12)$$

The mass is still set to zero as only one topo-cluster is used. The much more complex examples of TCCs $\textcircled{6}$ and $\textcircled{7}$ will not be detailed here.

An important detail of the TCC regarding pileup mitigation is that while TCCs cannot be seeded from tracks originating from pileup vertices, these tracks can be used in the energy sharing procedure. This means that TCCs built from a track is necessarily originating from the HS primary vertex but its energy scale takes into account energy overlap from pileup interactions.

From the TCC algorithm a collection of TCC objects is obtained that are either made of:

- one track originating from the HS primary vertex and one or multiple topo-clusters (referred to as *combined TCCs*),
- one topo-cluster not matched to any track (referred to as *neutral TCCs*),
- one track from the HS primary vertex not matched to any topo-cluster (referred to as *charged TCCs*).

Similarly to the PFlow collection, the TCCs are a reconstructed representation of events with no double energy counting between tracks and topo-clusters. Jets build from TCCs have great performances in terms of energy and angular resolution at high energy and can take advantage of the excellent angular separation between their constituents to study their internal structures (substructure).

Unified Flow Objects Both PFlow and TCC developments in jet reconstruction led to significant improvements in jet reconstruction. However, none of these two types of input constituents seems optimal for all jet metrics and for a wide range of energy. PFOs, as discussed above, are the best in the low- p_T range, improving correspondence between particles and reconstructed objects (tracks or topo-clusters). In contrast, at higher- p_T or in dense environment, the p_T resolution and angular separation power of the tracker are degraded, leading to degraded performance of the PFOs. On the other hand, the TCCs are the best at high- p_T , being able to resolve angular separations between several particles in dense environment. This allows improvement in jet internal structure measurements and thus in tagging performance. But at lower energy, the cluster splitting procedure is more prone to incorrect energy distribution between the different TCCs, leading to worse overall performance compared to PFlow. Moreover, TCCs show greater pileup instabilities at low p_T as the cluster splitting does not exclude pileup tracks. This motivated the ATLAS Collaboration to develop a third jet constituent reconstruction algorithm that would combine the best components of both PFOs and TCCs. This algorithm is (subsequently) called *Unified Flow Objects* (UFO) [82]. The UFO reconstruction workflow is schematised in Fig. 5.10.

First, the PFlow reconstruction algorithm is applied to obtain the charged and neutral PFOs. Any charged PFOs that originates from a pileup vertex is removed. The charged PFOs are separated in two sets:

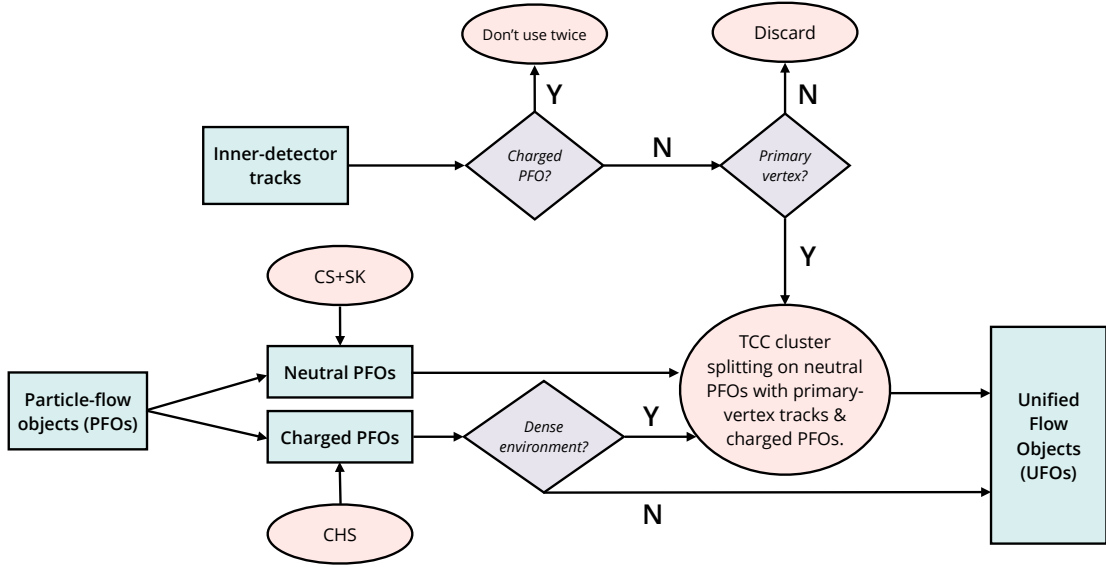


Figure 5.10: The Unified Flow Objects algorithm [82]. The CS+SK and CHS procedures are pileup mitigation techniques which will be detailed in the next section.

- charged PFOs with subtraction which are directly retained as UFOs,
- charged PFOs without subtraction because of dense environments.

Second, the TCC algorithm is applied on:

- tracks from the HS primary vertex that were not used for PFlow reconstruction ($p_T > 100$ GeV),
- neutral PFOs (isolated topo-clusters or remnants of the PFlow subtraction but also topo-clusters originating from high- p_T tracks),
- unsubtracted charged PFOs.

The obtained collection of TCCs and the subtracted charged PFOs are retained as UFOs. This double procedure allows to benefit from the PFlow algorithm performance at low p_T through the charged PFOs and the TCC algorithm at high- p_T through the reconstructed TCCs from unused high- p_T tracks. Jets reconstructed from UFOs show either similar performances as PFlow or TCC jets (in their respective energy range) or improved performance. It is expected that the UFO constituents will be use broadly throughout Run 3 for both small- R and large- R jet reconstruction.

5.4.3.3 True constituents

Finally, in MC simulations, the generated particles can also be used directly as jet constituents. The particles considered are the stable particles obtained after the simulation of the hadronisation processes by the generator, before the simulation of the detector response, at the exception of neutrinos and muons⁸. Jets built from true constituents are called true/truth jets in contrast with reconstructed jets built from

⁸ Neutrinos do not interact with detector and muons do not deposit significant enough energy in the calorimeters.

detector signals (e.g. PFOs, TCCs, UFOs). By geometrically matching true and reconstructed jets, it is then possible to compute a measure of the detector response in energy, p_T or mass by taking their ratios between true and reconstructed jet:

$$r_E = \frac{E_{\text{reco}}}{E_{\text{true}}}, r_m = \frac{m_{\text{reco}}}{m_{\text{true}}} \quad (5.13)$$

where the subscripts $_{\text{reco}}$ and $_{\text{true}}$ refer respectively to the reconstructed and true jet variables. The energy and mass responses r_E and r_m are key variables for jet calibration as it will be discussed briefly in Sec. 5.4.5 and thoroughly in Sec. 6.1.

5.4.4 Pileup mitigation and grooming techniques

Techniques for subtracting pileup or underlying event contributions from the particles originating from the HS primary vertex are available both at the jet constituent level and jet-level. These techniques generally rely on either removing tracks or topo-clusters before the constituent building or removing constituents before or after jet reconstruction.

5.4.4.1 Charged Hadron Subtraction

The most basic pileup mitigation technique is called Charged Hadron Subtraction (CHS) and consists in removing tracks or charged constituents which originate from pileup vertices. It can be applied during the constituent building or during jet reconstruction. In particular, this procedure is applied in the TCC and UFO algorithms and during PFlow jet reconstruction. The CHS allows in a simple way to take advantage of the track impact parameters at the perigee to only select the ones with d_0 and z_0 consistent with the HS primary vertex position.

5.4.4.2 Neutral constituent subtraction

More sophisticated techniques based on a pileup energy density estimation are also used on top of the UFO algorithm to remove neutral constituents.

The first one is called Constituent Subtraction (CS) [83] and consists in a local subtraction of pileup at the level of individual constituents. For UFO building, this procedure is applied to the neutral PFOs before the TCC algorithm. The CS involves the combination of constituent kinematics with soft "negative" particles that are added to the event to balance the pileup contributions. This procedure is derived from the pileup energy density contribution ρ estimated on an event-by-event basis. The pileup energy density is defined as the median of $p_T^{\text{jet}}/A^{\text{jet}}$ distribution of the $R = 0.4$ jets reconstructed with the k_t algorithm, where A^{jet} is the jet angular area. From this estimation, massless soft particles with transverse momentum p_T^g called ghosts are overlaid uniformly in the $\phi - \eta$ plane of the event such that:

$$p_T^g = A^g \cdot \rho$$

A^g is the area of the ghosts set to 0.01. p_T^g is the expected pileup contribution in an area $\Delta\eta \times \Delta\phi = 0.1 \times 0.1 = 0.01$ and the total p_T^g sum of all the ghosts is then equal to the expected average pileup

contribution. Therefore, these ghost particles defined as pileup only contributions can be used as "negative" particles that need to be subtracted to the jet constituents in order to mitigate pileup contributions. This subtraction is done through an iterative procedure that defines the specific amount of energy (p_T) to be subtracted to each constituent. First, the angular distance $\Delta R_{i,k}$ between each constituent i and ghost k is computed, and the pairs are ordered in ascending $\Delta R_{i,k}$. Then, proceeding iteratively through the pairs the algorithm modifies the p_T of each constituent and ghost with the following procedure:

$$\begin{aligned} \text{If } p_{T,i} \geq p_{T,k} : \quad & p_{T,i} \rightarrow p_{T,i} - p_{T,k}, \\ & p_{T,k} \rightarrow 0; \\ \text{otherwise:} \quad & p_{T,k} \rightarrow p_{T,k} - p_{T,i}, \\ & p_{T,i} \rightarrow 0. \end{aligned}$$

This procedure stops when $\Delta R_{i,k} > \Delta R_{\max}$. $\Delta R_{\max}$ is set to 0.25 to guarantee that only neighbouring ghosts to a constituent are used for p_T corrections. At the end of the CS algorithm, all remaining ghosts are removed and if some constituents p_T are completely subtracted they are also removed. The remaining corrected constituents are then available for the UFO algorithm.

The second neutral constituent subtraction applied in the UFO building is called SoftKiller (SK) [84]. It consists in applying a p_T cut on the constituents as most pileup contributions are soft emissions. The choice of the cut-off is made in each event so that after the constituent removal $\rho \simeq 0$. In practice, the event is divided in $\phi - \eta$ patches of length scale $l = 0.6$ and the p_T cut-off is chosen as the lowest cut for which half of the patches are empty after the removal. As ρ is defined as a median of jets p_T density, removing half of the patches by the SK procedure will lead to $\rho \simeq 0$ for each event.

The full UFO building procedure uses both the CS and SK procedures (generally referred as CSSK) in this specific order. The CS is first applied to correct the p_T of the constituents for pileup pollution in the constituents themselves. The SK is then applied to determine the constituents that are too soft and most likely originating from pileup contributions.

5.4.4.3 Jet grooming algorithms

Grooming algorithms are applied after the jet reconstruction, directly to the jets themselves to remove soft and/or wide-angle emissions present in the jets. These soft and wide angle emissions are typical of pileup and underlying event contributions and need to be removed as much as possible while keeping the hard emissions constituting the jets. These emissions are the most problematic for large- R ($R = 1.0$) jets which have an important area and are thus more sensible to these soft contributions. The grooming algorithms are thus only applied to large- R jets and aim at improving the resolution of the jet energy, mass and substructure measurements. The substructure variables are crucial for jet tagging but they can be largely sensible to soft and wide-angle emissions.

Historically, ATLAS used the Trimming algorithm [85] which aims at excluding regions of the jets with low energy density that most likely originate from soft and wide-angle emissions. The trimming algorithm consists in reclustering the jet constituents in smaller sub-jets of specific radius R_{sub} and removing the sub-jets which have a p_T^{sub} smaller than a specific fraction of the ungroomed jet p_T, f_{cut} . The constituents

of the removed sub-jets are discarded and the remaining ones form the groomed / trimmed jet. The usual choice for the parameters R_{sub} and f_{cut} are: $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.05$.

However, a new grooming algorithm is now used for UFO jets called Soft Drop (SD) [86] which shows better performance in terms of mass resolution and tagging ability [82]. In this algorithm, the jet constituents are reclustered using the C/A algorithm to create an angle-ordered jet clustering sequence. The clustering sequence is then unravelled in reverse (i.e. declustering), starting from the widest-angle constituents. At each clustering step p_T^1 and p_T^2 are defined, which are the p_T of the two branches resulting from the declustering splitting and ΔR^{12} , the angular separation between the two branches. Then, the following condition is tested:

$$\frac{\min(p_T^1, p_T^2)}{p_T^1 + p_T^2} < z_{\text{cut}} \left(\frac{\Delta R^{12}}{R} \right)^\beta$$

where R is the jet radius, z_{cut} and β are the SD parameters that control the amount of soft and wide-angle emission that is removed. If the clustering step and the two splitted branches fail this condition, the lower- p_T branch is removed and the declustering continues only with the higher- p_T branch. The procedure stops at the first clustering step that satisfies the condition. The remaining constituents form the groomed / soft-dropped jet. The usual choice for the SD parameters z_{cut} and β are: $z_{\text{cut}} = 0.1$ and $\beta = 1$.

5.4.5 Jet calibration

Jets built from reconstructed objects correspond to raw detector-level measurements and are thus subject to fluctuations in the interactions between the particles and the detector. This leads in general to kinematic differences between reconstructed and truth jets, embodied by a jet energy, mass or transverse momentum response (as defined in Eq. 5.13) different from one. Moreover, these interaction fluctuations are stochastic effects in the development of calorimeter showers. To each emitted true jet then correspond a complete distribution of possible responses. The calibration applied to jets aims, therefore, at correcting on average the reconstructed jet four-momentum to that of truth jets. This calibration is applied in multiple steps, each one consisting in a correction to the jet four-momentum, either scaling the jet p_T , energy or mass (or combination). Ref. [80] and more recently [87] present these different steps for small- R jet calibration which are schematised in Fig. 5.11. The large- R jet calibration does not correct for pileup effects (as grooming is applied before) and does not apply the global sequential calibration that is exclusive to small- R jets. However, the most important corrections which are the absolute MC-based and in-situ calibrations are common to both small- R and large- R jet calibration. A summary of the ATLAS jet calibration chain will be presented in this section.

5.4.5.1 Simulation-based calibration

The main calibration corrections applied to jets are derived from MC simulations, with the study of the jet kinematic responses (energy, mass or p_T) and their dependencies on the energy and environment (pileup) of the jets. The MC samples used are simulations of dijet (two high-energy quarks or gluons emitted back-to-back) and multijet (multi-partons high-energy emission) events produced in a large range of jet energy. These samples cover well the phase space of light QCD jets which constitute the majority of

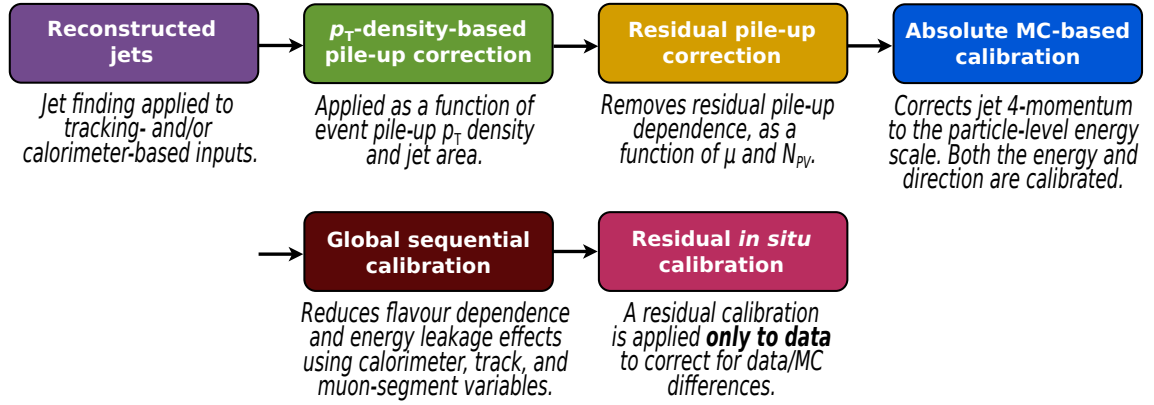


Figure 5.11: The small- R jet calibration chain of ATLAS [80].

jet production at the LHC. All the samples are overlaid with pileup interactions to simulate the LHC conditions.

Pileup corrections There are two corrections applied for pileup, one based on the pileup density ρ defined earlier with area subtraction and the other called residual pileup correction mitigating the dependency on N_{PV} and μ . The pileup density correction subtracts the expected energy contributions from pileup based on the pileup density ρ and the jet area. ρ is defined in the same way as for the CS algorithm, from the median of the $p_T^{\text{jet}}/A^{\text{jet}}$ distribution of the $R = 0.4$ jets reconstructed with the k_t algorithm. This correction assumes that the pileup contributions are uniform over a given jet area such that, for a jet with angular area A , the pileup contribution to the jet reconstructed p_T can be estimated as $\rho \times A$. The pileup density correction is then defined as:

$$p_T^{\text{area}} = p_T^{\text{reco}} - \rho \times A$$

In the specific case where the calibrated jet collection uses constituents after the SK pileup mitigation technique (usually UFO jets), the pileup density correction is not applied as it would be redundant.

To further reduce the impact of pileup, a second correction called residual pileup correction is applied. It aims at reducing the dependency of the jet four-momentum measurement on pileup. This correction is based on N_{PV} , μ , p_T^{reco} and η . N_{PV} and μ , quantify the pileup intensity but are sensitive to different type of pileup. N_{PV} being based on the primary vertices reconstruction in one event is sensitive to in-time pileup, while μ is sensitive to out-of-time pileup as it is measured from the average amounts of pileup over multiple bunch crossings. Therefore, a higher N_{PV} leads to an increase in jet energy, while a higher μ tends to lead to a decreased jet energy due to the LAr calorimeter specific pulse shape (see Fig. 5.7). These two behaviours motivate the use of N_{PV} and μ to decrease the pileup dependency. The additional residual pileup correction is defined as:

$$p_T^{\text{residual}} = p_T^{\text{area}} - (\partial p_T / \partial N_{PV}) \times (N_{PV} - 1) - (\partial p_T / \partial \mu) \times \mu$$

where $\partial p_T / \partial N_{PV}$ and $\partial p_T / \partial \mu$ measure the dependency on N_{PV} and μ of the p_T fluctuation between reconstructed and true jet. This correction removes the p_T difference between p_T^{area} and p_T^{true} for a given N_{PV} and μ such that the jet p_T scale matches that of the true jet. $\partial p_T / \partial N_{PV}$ and $\partial p_T / \partial \mu$ are determined with a sequential method based on $\Delta p_T^{\text{area-true}} = p_T^{\text{area}} - p_T^{\text{true}}$ measured in multiple bins. For $\partial p_T / \partial N_{PV}$, the $\Delta p_T^{\text{area-true}}$ dependency on N_{PV} is defined as the slope of a linear fit of $\Delta p_T^{\text{area-true}}(N_{PV})$ in bins of μ , p_T^{true} and η . This slope gives $\partial p_T / \partial N_{PV}$ per μ , p_T^{true} and η bins. To remove the μ dependency, the average over the μ bins is taken for each p_T^{true} and η bins. Then to remove the dependency on p_T^{true} , a logarithmic fit of $\partial p_T / \partial N_{PV}$ over p_T^{true} in each η bin is conducted. And the fit value at $p_T^{\text{true}} = 25$ GeV is taken as the nominal $\partial p_T / \partial N_{PV}$ correction per η bin. The choice to take only the fit value at $p_T^{\text{true}} = 25$ GeV comes from the fact that pileup effects are most significant at low jet p_T . Finally, to remove the discontinuities coming from the η binning, a last linear fit of $\partial p_T / \partial N_{PV}$ values over $|\eta|$ is applied. $\partial p_T / \partial N_{PV}$ is therefore the slope of $\Delta p_T^{\text{area-true}}(N_{PV})$ averaged over μ and for $p_T^{\text{true}} = 25$ GeV, as a function of $|\eta|$. For $\partial p_T / \partial \mu$, it is obtained in the same way by switching the places of N_{PV} and μ .

Both pileup corrections can be rounded up in one equation:

$$p_T^{\text{corr}} = p_T^{\text{reco}} - \rho \times A - \alpha \times (N_{PV} - 1) - \beta \times \mu \quad (5.14)$$

where α and β are respectively $\partial p_T / \partial N_{PV}$ and $\partial p_T / \partial \mu$. The ratio $p_T^{\text{corr}} / p_T^{\text{reco}}$ is used as the first correction factor to the jet four-momentum.

More recently, a more precise residual correction was developed to take into account the correlation between N_{PV} and μ , and the pileup contribution dependency on the jet p_T . This new correction is the 3D residual pileup correction described in [87] and is a function of $(N_{PV}, \mu, p_T^{\text{area}})$ (instead of individual N_{PV} and μ corrections). The correction is determined similarly as the first residual method.

Absolute MC-based calibration This part constitutes the main correction factor applied to the jet four-momentum and is based on the study of the jet energy (and also mass for large- R jets) response. As it is one of the main focus of this thesis, this part will be described in details in a dedicated section of Chapter 6. The absolute MC-based calibration corrects the energy, mass and η of jets so that it matches on average the scale of the true jets.

Global sequential calibration The main dependency of the jet response being on the energy and η , the absolute calibration corrects globally the jet response, with the central value of response distributions shifted to one within 1% on most of the jet phase space. However, other jet characteristic can contribute to the jet response such as the jet composition (fraction of charged / neutral constituents) or the energy deposit distribution in the different calorimeter layers. These contributions are not taken into account by the absolute calibration, hence jets with similar energy and η but different individual characteristics can have slightly different energy responses. Moreover, these jet individual characteristics can depend widely on the jet nature, quark- or gluon-initiated, and on the jet flavour. Such dependencies imply that datasets with no a priori knowledge about the jet flavour composition will present variations in global jet responses and wider energy resolutions resulting in larger uncertainties. It is then important to decrease such dependencies by applying correction factors as function of the jet individual characteristics so that varied jet samples will have similar energy scales. This is done by the global sequential calibration (GSC) and is exclusive to small- R jet calibration as the performance requirement on the jet response of small- R are tighter.

The GSC uses six jet observables that describe the shapes of the jet calorimeter showers, tracking characteristics and muon activity linked to the jet, and derives six independent corrections applied sequentially. These observables are:

- the fraction of the jet p_T carried by tracks associated to the jets,
- the fraction of the jet energy measured in the first layer of the HCal,
- the fraction of the jet energy measured in the third layer of the EMCal,
- the number of tracks with $p_T > 1$ GeV associated to the jet,
- the average p_T -weighted angular distance ΔR between the jet axis and all tracks with $p_T > 1$ GeV associated to the jet, $\langle p_T^{\text{trk}} \times \Delta R(\text{jet}, \text{trk}) \rangle$,
- the number of muon tracks associated to the jet.

For each GSC observable, X_{GSC} , the jet p_T ⁹ responses are first gathered in bins of $|\eta|$, p_T^{true} and X_{GSC} . The p_T response distribution is fitted in each bin with a gaussian distribution to obtain the mode of the distribution $R(X_{GSC}, p_T^{\text{true}}, |\eta|)$ (as for the absolute calibration, only on average corrections are derived). Then for each bin of X_{GSC} , a numerical inversion procedure [88] is used to remove the dependence of the p_T response fits on p_T^{true} and replace it by the jet reconstructed and corrected p_T . The resulting responses $R(X_{GSC}, p_T, |\eta|)$ are then smoothed simultaneously in p_T and X_{GSC} for each $|\eta|$ bin to get a continuous correction factor. This correction factor is ultimately defined as:

$$C(X_{GSC}, p_T, |\eta|) = \frac{1}{R(X_{GSC}, p_T, |\eta|)}$$

The numerical inversion method is similarly used to obtain the absolute MC-based calibration factors and will be precisely detailed in Chapter 6.

The GSC procedure is limited by the need to have uncorrelated observables X_{GSC} , since correlations would create interferences in the sequential corrections. This limitation, therefore, reduces widely the set of possible observables and thus the set of corrections. A more advanced method was thus developed using neural network techniques [87], which naturally take correlations into account. With this method, the GSC correction is performed with a single network taking multiple observables as input.

5.4.5.2 Data-based calibration: In-situ corrections

The MC-based calibration allows to correct the jet energy (and mass for large- R jets) scale to that of the particle level in MC simulations. A second calibration step is then necessary to correct for differences between the MC simulations and the real data. These differences come from mismodelling in the pp collisions generation and in the simulation of the detector material response. To account and correct for these, a data-based calibration called *in-situ* calibration is applied on top of the simulation-based calibration for jets in real data only. This calibration is based on the comparison of the jet response in MC simulation and in data. Here, the jet response is measured as the ratio between the MC-based calibrated jet p_T and the p_T of a well-calibrated reference object or system:

⁹ After the pileup and absolute calibrations.

$$r = \frac{p_{\text{T}}^{\text{cal,jet}}}{p_{\text{T}}^{\text{ref}}}$$

By performing this measure in events with a jet recoiling from a reference object or system, the overall event p_{T} balance requires r to be equal to one if the jet is well calibrated. Therefore, r measures the inefficiencies in the jet MC-calibration. This response is measured in bins of $p_{\text{T}}^{\text{ref}}$, and the average (or mode) is taken by fitting the response distribution with a gaussian, to obtain the in-situ average response:

$$\mathcal{R}_{\text{in-situ}} = \left\langle \frac{p_{\text{T}}^{\text{cal,jet}}}{p_{\text{T}}^{\text{ref}}} \right\rangle$$

This average response measured in data could be used directly as a calibration factor to account for the data-MC differences as one could expect that if $\mathcal{R}_{\text{in-situ}}$ is not equal to one it is because of the data-MC inconsistencies. However, $\mathcal{R}_{\text{in-situ}}$ is also sensitive to additional radiation jets in the event or additional / leaked energy during jet reconstruction that would bias the measure of $\mathcal{R}_{\text{in-situ}}$. For that reason, the double ratio between $\mathcal{R}_{\text{in-situ}}$ measured in data and MC is defined:

$$C = \frac{\mathcal{R}_{\text{in-situ}}^{\text{data}}}{\mathcal{R}_{\text{in-situ}}^{\text{MC}}}$$

This double ratio is much more robust to these secondary effects as the same effects would appear in both the data and MC¹⁰. The final in-situ calibration factor is then obtained by a numerical inversion of C from a function of $p_{\text{T}}^{\text{ref}}$ to a function of the MC-calibrated jet p_{T} .

There are three in-situ stages, each based on different reference object or system. The first one is the η intercalibration which uses the p_{T} balance in dijet events to correct the jet energy scale of forward jet ($0.8 < |\eta| < 4.5$) to match the scale of better calibrated central jets. This is done because forward regions of the detector are usually less understood due to the complex detector structure (transition between barrel and end-cap calorimeters, end of the tracker). The second stage uses a well calibrated Z boson or photon to balance a hadronic recoil reconstructed as a jet. The last stage uses a multi jet balance with a system of well calibrated low- p_{T} jets balancing a single high- p_{T} jet. The low- p_{T} jets are calibrated up to the Z/γ +jet in-situ calibration and the high- p_{T} jet is left at the η intercalibration stage. This last step is necessary for high- p_{T} jets as the Z/γ +jet corrections is limited in p_{T} range due to low statistics at higher p_{T} .

In addition, for large- R jets an *in-situ* jet mass correction is applied following the *in-situ* energy correction. The methods for deriving an *in-situ* jet mass correction are described in Ref. [89] but are not detailed here.

¹⁰ At the condition that these secondary effects are well simulated of course.

CHAPTER 6

SIMULTANEOUS CALIBRATION OF THE ENERGY AND MASS OF LARGE-RADIUS JETS USING A DEEP NEURAL NETWORK

Outline of This Chapter

6.1	The standard jet calibration in ATLAS	76
6.2	Samples, reconstruction and selections	79
6.3	Introduction to Deep Neural Networks	81
6.4	Deep Neural Network calibration methodology	83
6.5	Performances and validation of the calibration	91
6.6	Limitations and possible solutions	106
6.7	Conclusion	110

The first main project of my thesis was the development of a new calibration method for large- R jets using machine learning techniques in collaboration with Pierre-Antoine Delsart (CNRS-IN2P3, LPSC, Grenoble). As presented in Chapter 5, jets reconstructed from raw detector measurements (tracks, calorimeter clusters) have to follow a series of corrections to precisely measure their angular positions and energies. In this procedure, the main energy and mass corrections are evaluated numerically from simulated collision events during the absolute MC-based calibration step. These corrections are evaluated and applied in two distinct successive steps, in bins of a limited (usually two or three) number of characteristic variables. However, calibrating both highly correlated energy and mass while exploiting all jet structure variables and their correlations in a single correction step is possible through approaches based on deep neural networks (DNN). In that way, I took part in a full implementation of a DNN simultaneous energy and mass calibration for large- R jets developed within the ATLAS Collaboration using simulated collision events, overcoming several difficulties that are specific to the calibration problem. It is planned that this new calibration method will at term replace the current absolute MC-based calibration of large- R jets for all upcoming Run 2 and Run 3 analyses. This work ultimately led to the writing of a Collaboration paper [90] submitted and published in the Machine Learning: Science and Technology (MLST) IOP journal and for which I am one of the two main authors. Therefore, this Chapter presents this implementation, which is largely based on the published paper, adding and adapting sections were necessary.

First, Section 6.1 reviews the current absolute MC-based calibration procedure used in ATLAS (briefly introduced in Section 5.4.5) and motivates the new DNN approach. Section 6.2 describes the MC simulation

datasets used, and the jet reconstruction and selection procedures. A short introduction to DNNs is given in Section 6.3. Section 6.4 details the specific methodology needed to design and properly train the DNN. Then Section 6.5 discusses the validation and performance of the resulting DNN calibration, comparing it with the current numerical method. Finally, Section 6.6 identifies and discusses limitations in this approach as well as some possible solutions.

6.1 The standard jet calibration in ATLAS

The absolute MC-based calibration for large- R jets corrects the jet energy and mass successively. It is the main correction applied to large- R jets and the only one based on simulation. These corrections are derived sequentially from studies of the *individual response* of the quantities of interest E , the jet energy, and m , the jet mass. The following discussion will present the calibration method and corresponding equations for the jet energy E , but all of these stand identically for the mass m . The individual jet energy response is defined by

$$r_E = \frac{E_{\text{reco}}}{E_{\text{true}}}$$

where E_{true} is calculated from the true jets and E_{reco} from the jets reconstructed after detector simulation. Because of the stochastic nature of both the QCD shower and its interactions with the detector, true jets that have the same parameters (such as energy, angular position, mass, . . .) can generate distributions of possible r_E . Individual response distributions depend on these parameters and are usually unimodal: their most probable value, their mode R_E , defined as

$$R_E = \text{mode}(\{r_E\})$$

is called the *average response*, JES (or JMS in the equivalent case for jet mass), or simply *response* when the context is non-ambiguous. The mode, as opposed to the mean or the median, is chosen as the definition of such a central value to avoid ambiguities or biases that can arise when the distributions are asymmetric. The width of the distribution around the mode is used to define the resolution in E . In this Chapter, the chosen definition is the inter quantile range (IQR) at 68%¹ divided by the mode:

$$\sigma_E = \frac{\text{IQR}_{68\%}(\{r_E\})}{R_E} \quad (6.1)$$

Because of experimental limitations introduced in Sec. 5.4.5, the average response may differ from unity and the objective of the calibration, as illustrated in Fig. 6.1, is to find the relevant corrections such that $R_E = 1$ for all populations of true jets (which formalizes the idea that true and reco scales are *on average* identical), with σ_E as small as possible.

The individual response distributions can be parameterised by a set of parameters describing the true jet (here taken to be the energy, pseudorapidity and mass) but also by inter-dependent parameters of the reconstructed jets such as the energy, pseudorapidity, mass, fractions of energy deposited in the various layers of the calorimeter, fraction of the momentum of the associated tracks, and others. Formally, it is then possible to model the distribution as a probability distribution function $P(r_E | \vec{x})$ where $\vec{x}$ is the list of parameters that can be separated into the various true and reconstructed parameters $\vec{x} = (\vec{x}_{\text{true}}, \vec{x}_{\text{reco}})$. The average response for true jets with parameters $\vec{x}_{\text{true}}$ is then given by

$$R_E(\vec{x}_{\text{true}}) = \text{mode}(P(r_E | \vec{x}_{\text{true}})). \quad (6.2)$$

¹ IQR instead of the standard deviation is chosen here as the shape of the response distribution does not always follow a Gaussian distribution, particularly for the jet mass.

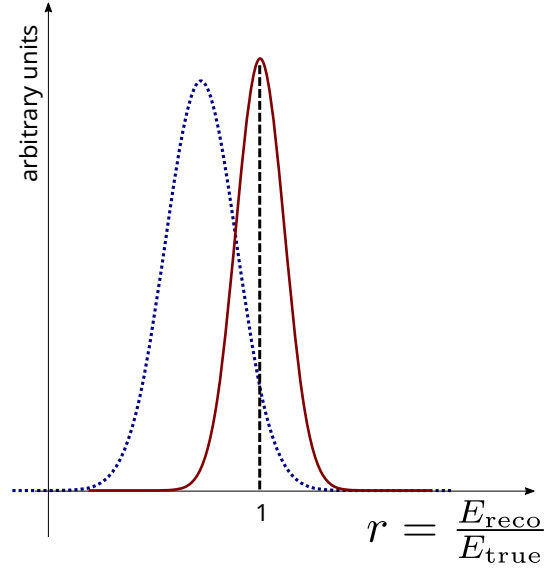


Figure 6.1: Schematic distribution of uncalibrated (blue) and calibrated (red) energy response distributions.

The aim of the calibration is then to find a function of the reconstructed parameters $C(\vec{x}_{\text{reco}})$ defining the calibrated quantity $E_{\text{calib}} = C(\vec{x}_{\text{reco}})E_{\text{reco}}$ such that the most probable value becomes

$$R_{E_{\text{calib}}}(\vec{x}_{\text{true}}) = \text{mode}(P(r_E \times C(\vec{x}_{\text{reco}}) \mid \vec{x}_{\text{true}})) = 1$$

where $P(r_E \times C(\vec{x}_{\text{reco}}))$ is the probability distribution function obtained from $P(r_E)$ by changing $r_E \rightarrow r_{E_{\text{calib}}} = r_E \times C(\vec{x}_{\text{reco}})$ and the mode is taken over all $\vec{x}_{\text{reco}}$.

In general this mathematical problem is difficult. It has an exact solution in simple cases where the response depends on only one parameter, the quantity E itself, and when $R_E(E_{\text{true}})E_{\text{true}}$ is an affine function [88]. In this case the solution is to choose the function C such as

$$\forall x \in \mathbb{R}, \quad C(xR_E(x)) = \frac{1}{R_E(x)}. \quad (6.3)$$

This solution is called the *numerical inversion* (NI) because R_E (and thus the pair $(xR_E(x), \frac{1}{R_E(x)})$) can be obtained *numerically* from the simulated data and because $E_{\text{true}} R_E(E_{\text{true}})$ represents an *inversion* of the true quantity.² This method (referred to here as *standard calibration*) is employed by ATLAS to obtain sequentially a jet energy scale calibration and then a mass scale calibration. In practice, in the standard calibration, R_E is evaluated as a function of $(E_{\text{true}}, \eta_{\text{true}})$ (and m_{true} for mass calibration) by collecting individual responses in bins of these quantities. This function R_E is then converted into a correction factor function depending only on reconstructed quantities $(E_{\text{reco}}, \eta_{\text{reco}}$ and $m_{\text{reco}})$ according to the NI principles described above and further detailed in [80, 91]. This procedure gives satisfactory performance. The variables E , η and m are the principal dependencies of the JES and JMS and the use of true quantities

² For a given E_{true} , the most probable E_{reco} is by definition $E_{\text{true}} R_E(E_{\text{true}})$.

to form the response evaluation bins allows potential biases due to the energy or mass spectrum of the generated sample to be avoided when evaluating the average response.

As mentioned above, there is much more information available to describe the reconstructed hadronic showers than just the energy, pseudorapidity and mass. Using these additional variables can potentially help obtain a more precise JES and JMS and improve their resolution, including subsamples of jets such as quark-initiated or gluon-initiated jets. However, these variables can be correlated. Optimally exploiting all the correlations in the same way as performed in the standard calibration would require highly multidimensional bins, which is practically impossible. On the other hand the task of numerically modelling a multidimensional function can be successfully performed with modern DNN techniques. Furthermore, a key problem of the calibration evaluation, estimating the mode of distributions, is solvable with these techniques; a solution is discussed in detail in Section 6.4.1. A proof of concept of using DNN for JES and JMS calibration, including an implementation of the NI procedure, has already been presented in [92] and DNN techniques are also employed to perform the GSC step of the energy (but not the mass) calibration for small- R jets as discussed in Section 5.4.5. Building on this, a complete and operational solution and its performance are presented here. This new procedure extends several aspects of the DNN calibration but does not implement NI to simplify the method by avoiding the need to train two DNNs sequentially, one to model the function R as a function of x_{true} , the second the function C as in Eq. (6.3). Instead, it evaluates directly the uncalibrated responses in terms of reconstructed quantities, which means that the network evaluates

$$R_E(\vec{x}_{\text{reco}}) = \text{mode}(P(r_E | \vec{x}_{\text{reco}})) \quad (6.4)$$

instead of $\text{mode}(P(r_E | \vec{x}_{\text{true}}))$ as in Eq. (6.2). The calibration function is then chosen to be simply $1/R_E(\vec{x}_{\text{reco}})$. The price to pay for this simpler procedure is that the response $\text{mode}(P(r_E | \vec{x}_{\text{reco}}))$ a priori depends on the distribution of the simulated input samples, contrary to $\text{mode}(P(r_E | \vec{x}_{\text{true}}))$ used in the standard calibration, and this dependency can bias the resulting calibration even in the simple cases where the standard calibration would not be biased. However, because $\vec{x}_{\text{reco}}$ includes many variables relevant to the description of the hadronic shower, it can be expected that the response distributions at each point of the phase space are narrow enough so that the overall bias on the mode is negligible. Although this hypothesis is not testable in practice because of the high dimensionality, verifications are performed to ensure the input samples do not bias the performance (they are presented in Section 6.5.3).

6.2 Samples, reconstruction and selections

6.2.1 Event simulation

The training and validation of the DNN is based on MC simulations of pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV (Run 2 conditions). Various physics processes were simulated as described in Table 6.1 to train and test the DNN on different jet topologies.

Table 6.1: Simulated samples used for training and validation

Use	Process type	Generator	Number of jets passing the selections (in millions)
Training, validation	QCD dijet	PYTHIA 8.230	~ 270
Validation	W/Z	PYTHIA 8.230	~ 20
Validation	Top quark	PYTHIA 8.230	~ 15
Validation	Higgs Boson	PYTHIA 8.230	~ 15
Validation	QCD dijet	SHERPA 2.2.5	~ 30

Training is performed on the QCD dijet sample as this type of process is, by orders of magnitude, more probable at the LHC and enters as a background to many analyses. In a validation step, it is verified that the training generalises and produces a good calibration for other important physics processes such as those listed in Table 6.1. All samples (except the SHERPA one) were produced using PYTHIA 8.230 [60] with the NNPDF2.3LO [93] parton distribution function (PDF) set and the A14 set of tuned parameters [94]. The training sample containing jets originating from light quarks or gluons was obtained by generating dijets in slices of the jet transverse momentum, p_T , to sufficiently populate the kinematic region of interest (from around 200 GeV up to 6000 GeV). For the validation samples, physics processes modelling phenomena beyond the Standard Model are used to populate the high- p_T region:

- The W/Z -boson sample was obtained by generating the $W' \rightarrow WZ \rightarrow q\bar{q}q\bar{q}$ process with $m_{W'} = 2$ TeV.
- The top-quark sample was obtained by generating the $Z' \rightarrow t\bar{t} \rightarrow W^+b W^- \bar{b} \rightarrow bq\bar{q} \bar{b}q\bar{q}$ process with $m_{Z'} = 4$ TeV.
- The sample of Higgs bosons, which displays a broad Higgs boson p_T spectrum, was obtained by generating the production of Randall–Sundrum gravitons G^* in a benchmark model with a warped extra dimension [95] over a range of graviton masses from 300 to 6000 GeV, where the graviton decays according to $G^* \rightarrow HH \rightarrow b\bar{b}b\bar{b}$.

The SHERPA dijet sample was generated using the SHERPA 2.2.5 [61] generator. The matrix element calculation was included for the $2 \rightarrow 2$ process at leading order, and the default SHERPA parton shower [96] based on Catani–Seymour dipole factorisation was used for the showering with p_T ordering, using the CT14NNLO PDF set [97]. These samples used the dedicated SHERPA AHADIC model for hadronisation [98], based on cluster fragmentation ideas.

Finally, the effect of pileup was modelled by overlaying the simulated hard-scattering event with inelastic pp events generated with PYTHIA 8.186 [99] using the NNPDF2.3LO set of PDFs [93] and the A3 set of tuned parameters [100]. The number of overlaid events is representative of the amount of pileup present in the 2017–2018 ATLAS pp data sample (end of Run 2). However, the content in truth particles of

the datasets only includes particles generated by the hard-scatter event, such that the true jet reference is independent of pileup condition.

6.2.2 Jet reconstruction and selection

The reconstructed jets are build from the UFO constituents described in 5.4.3. They are clustered into jets with the anti- k_t jet algorithm with angular radius $R = 1.0$ to obtain large- R jets. Then, they undergo the CSSK and Soft Drop pileup mitigation procedures described in 5.4.4, the Soft Drop procedure is employed with the default parameters $\beta = 1$ and $z_{\text{cut}} = 0.1$.

True jets are obtained by clustering with the same algorithm and angular radius the stable particles produced by the generators (except neutrinos and muons). They similarly undergo the Soft Drop procedure to mitigate effects from the underlying event.

To obtain training samples that are not biased by experimental effects such as pileup, noise or misidentifications, some requirements are imposed on the jets. Reconstructed jets are selected if they are matched to a true jet and isolated as described in Ref. [80]. A reconstructed and a true jet are matched if they are within an angular distance $\Delta R = 0.3$. Two isolation requirements are then applied; reconstructed jets are rejected if a true jet other than the matched one is found within $\Delta R = 2.5$ or if another reconstructed jet is found within $\Delta R = 1.5$. Finally, both the reconstructed and true jets are required to have $p_T > 50 \text{ GeV}$ and, in an event, only the two highest- p_T jets satisfying these criteria are considered. The selection results in about 270 million jets being available for the DNN training sample.

6.3 Introduction to Deep Neural Networks

A DNN can be seen as a succession of analytic operations applied to a vector of input features $\vec{x}$ and that outputs one or more numbers, $\vec{y}_{\text{pred}} = f(\vec{x}, \vec{\theta})$, as a function of the input and the parameters $\vec{\theta}$ entering the definitions of these operations. It is usually represented as in Figure 6.2 with several successive layers. The first one (in orange) is the input vector $\vec{x}$. The following ones are called hidden layers and define the DNN architecture. Each one is defined by a number of neurons N_i (blue points), a matrix of weights W_i of size $N_i \times N_{i-1}$, a vector of biases $\vec{b}_i$ of size N_i and a non-linear activation function ϕ_i . It takes as input the previous layer $\vec{h}_{i-1}$ and apply the following operation :

$$\vec{h}_i = \phi_i(W_i \vec{h}_{i-1} + \vec{b}_i)$$

$\vec{\theta}$ then represents the set of W_i matrices and $\vec{b}_i$ biases' vectors,

$$\vec{\theta} = \{W_i, \vec{b}_i\}$$

All these operations lead to a complex non-linear transformation of $\vec{x}$ that can approximate any bounded and regular function of $\mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ [101].

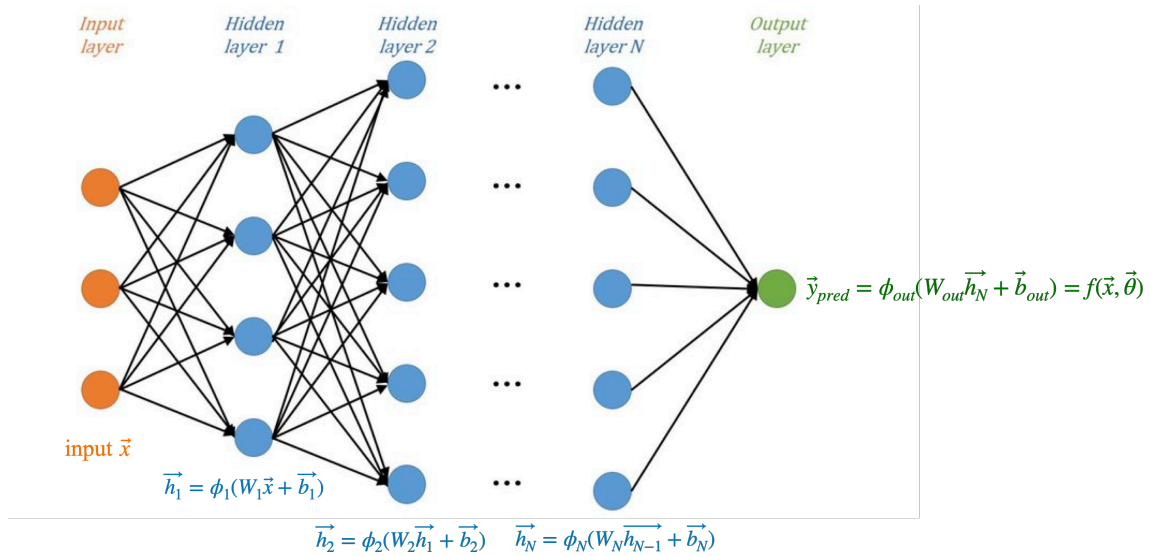


Figure 6.2: Schematic of the architecture of a Deep Neural Network.

Initially all DNN parameters ($\vec{\theta}$) are assigned randomly. Hence, a newly initialised DNN will predict nothing by just applying a random transformation to the input. In order to obtain the desired output, the DNN needs to be trained. This is done by tuning $\vec{\theta}$ in order to improve the DNN's predictions $\vec{y}_{\text{pred}}$ knowing what are the expected outputs, the targets $\vec{y}_{\text{true}}$. The training uses a Loss function $\mathcal{L}$:

$$\mathcal{L}(\vec{y}_{\text{pred}}, \vec{y}_{\text{true}})$$

This function is usually chosen to be minimal for $\vec{y}_{\text{pred}} = \vec{y}_{\text{true}}$ and serves as a quantifier of the difference between predictions and targets. The training principle is to minimise the average of $\mathcal{L}$ over the training sample, in order to minimise the difference between $\vec{y}_{\text{pred}}$ and $\vec{y}_{\text{true}}$ ¹ and improve the prediction. The minimisation uses back-propagation techniques to evaluate the gradient $\frac{\partial \mathcal{L}}{\partial \theta}$ [102] and algorithms similar to the Stochastic Gradient Descent (SGD) to update $\vec{\theta}$. The SGD formula to update a weight is given by :

$$\theta_i^{\text{new}} = \theta_i^{\text{old}} - \epsilon \frac{\partial \mathcal{L}(\vec{x}, \vec{y}_{\text{true}}, \vec{\theta})}{\partial \theta_i}(\theta^{\text{old}})$$

where ϵ is the learning rate, $\theta_i^{\text{new/old}}$ the updated/previous weight i , θ^{old} the previous set of weights and $\mathcal{L}(\vec{x}, \vec{y}_{\text{true}}, \vec{\theta})$ the loss function evaluated for the input $\vec{x}$ and corresponding target $\vec{y}_{\text{true}}$. It involves the gradient of the loss with respect to all the DNN parameters θ_i evaluated with the previous set of weights.

The usual training procedure is resumed in figure 6.3. Training data containing inputs $\vec{x}$ and targets $\vec{y}_{\text{true}}$ is fed to the DNN. For each input, the DNN outputs predictions $\vec{y}_{\text{pred}}$ and the weights $\vec{\theta}$ are updated based on the loss for this input.

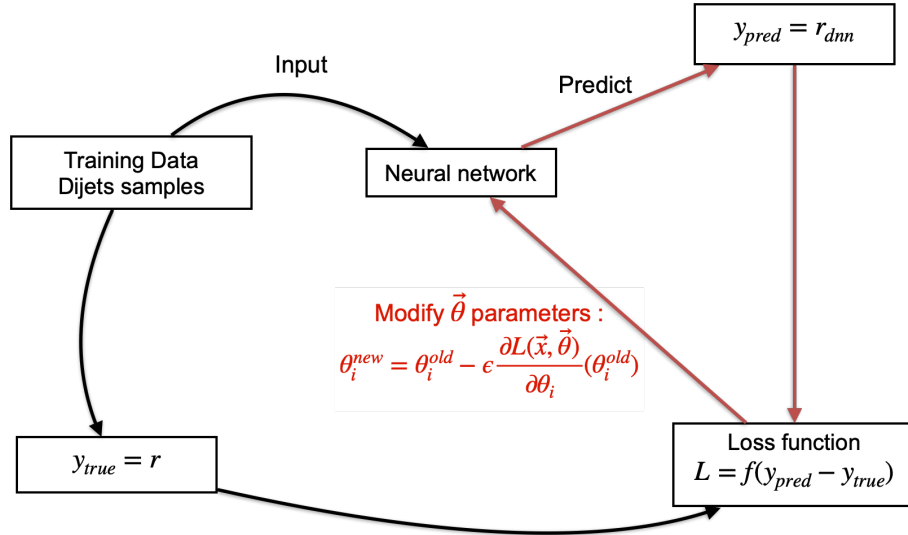


Figure 6.3: Schematic of the training procedure of a Deep Neural Network.

In practice, the DNN weights are not updated for each input as it would be computationally too expensive, on the contrary the updates are done at the end of what is called a batch. A batch is a sub-set of the training dataset compose of N inputs. For each input, predictions are made and the corresponding losses computed, then the mean over the N corresponding gradients is used to updates the DNN parameters. The training dataset can be used multiple times to gradually improve the predictions. Each complete iteration over the training dataset is called an epoch. For each epoch, the loss function as well as the batch size and other hyperparameters can be changed at convenience.

If the DNN setup (architecture, loss, training procedure) has been well-designed, at the end of the training the DNN should be able to predict the right outputs whether it be for the training data or for completely new data (this is called generalisation).

¹ The DNN training is similar to a minimisation problem.

6.4 Deep Neural Network calibration methodology

The simultaneous calibration of the jet energy and mass is performed using the calculations of a single DNN taking as inputs reconstructed jet-level and event-level quantities. The DNN is set up to output two numbers corresponding to the energy and mass responses. However, as mentioned in the previous section, it is not possible to directly predict the exact individual E (or m) response for a given reconstructed jet as each true jet corresponds to a distribution of possible responses. The DNN predictions are thus interpreted as the most probable values of the possible responses for the input jets (that is, the modes of their distribution). Following Eq. (6.4), the DNN is trained so that

$$R_E^{\text{DNN}}(\vec{x}_{\text{reco}}) \simeq R_E(\vec{x}_{\text{reco}}) \equiv \text{mode}(P(r_E | \vec{x}_{\text{reco}})).$$

The calibrated energy (or mass) is then defined as

$$E_{\text{calib}} = \frac{E_{\text{reco}}}{R_E^{\text{DNN}}(\vec{x}_{\text{reco}})}.$$

Thus, the DNN output predictions are

$$\vec{y}_{\text{pred}} = (R_E^{\text{DNN}}, R_m^{\text{DNN}}).$$

During the training, the target quantities corresponding to the inputs $\vec{x}_{\text{reco}}$ are the individual responses discussed in Section 6.1:

$$\vec{y}_{\text{target}} = \vec{y}_{\text{true}} = (r_E, r_m) = \left(\frac{E_{\text{reco}}}{E_{\text{true}}}, \frac{m_{\text{reco}}}{m_{\text{true}}} \right).$$

These are not directly the desired *average* responses R_E and R_m and the necessity to predict distribution modes (and not means nor medians) requires a careful choice of the loss function. Beside these mathematical concerns, experimental difficulties related to the detector geometry – and especially the geometry of the calorimeters – strongly interfere in the calibration predictions for certain angular regions. This is accounted for by a dedicated processing of the jet angular position.

Finally, the mass calibration alone has specific difficulties. Firstly, the mass response distributions are generally wider and asymmetric, making it harder for the DNN to predict their mode. Secondly, large- R jets are mainly used to reconstruct heavy particles whose decays are collimated (top quarks, W , Z , or Higgs bosons) and for which the mass measurement is crucial. So the DNN performance needs to be optimal for these jets, even if their topology differs from standard QCD jets. This has motivated the use of a specific architecture (including a *residual connection*-like structure) and training procedure (with a dedicated mass-only step).

6.4.1 Loss function

The loss function $\mathcal{L}(\vec{y}_{\text{true}}, \vec{y}_{\text{pred}})$ is a crucial part of the DNN training as it defines what mathematical quantity the DNN will learn to predict. This is particularly true in the case presented here where the targets ($\vec{y}_{\text{true}}$) are realizations of a stochastic distribution. In this case, it is proven that minimising the usual mean square error loss ($\mathcal{L} = (\vec{y}_{\text{true}} - \vec{y}_{\text{pred}})^2$) would have $\vec{y}_{\text{pred}}$ converge to the mean of the distribution, while minimising the mean absolute error ($\mathcal{L} = |\vec{y}_{\text{true}} - \vec{y}_{\text{pred}}|$) would imply a convergence to the median [103]. Converging to the distribution mode, as is necessary for the jet calibration, would require taking a Dirac delta distribution as the loss $\mathcal{L} = \delta(\vec{y}_{\text{true}} - \vec{y}_{\text{pred}})$. This is not possible in practice and an approximate function is therefore necessary. One possibility is the Gaussian kernel as suggested in Ref. [103], another solution is inspired by mixture-density-network (MDN) techniques and is described below. Numerical tests show the MDN losses are less biased than the Gaussian kernel losses. They are also found to converge faster and allow more flexibility in the training procedure.

First, as the DNN produces two outputs $\vec{y}_{\text{pred}} = (R_E^{\text{DNN}}, R_m^{\text{DNN}})$, two independent loss functions are considered, the sum of them being the final minimised quantity:

$$\mathcal{L}_{\text{tot}} = \mathcal{L}_E + \mathcal{L}_m.$$

The MDN loss is constructed from the assumption that the response distribution of the DNN targets is a Gaussian distribution with central value μ and width σ .

$$P(r_E, (\mu, \sigma)) \propto \frac{e^{-(r_E - \mu)^2 / 2\sigma^2}}{\sigma}.$$

With the parameter of interest being μ (the mode of the distribution), an estimator $\hat{\mu}$ (and $\hat{\sigma}$) can be obtained from the maximum of the likelihood (LH) over a given sample of r_E

$$\text{LH}(\mu, \sigma) = \prod_{r_E} P(r_E, (\mu, \sigma)). \quad (6.5)$$

As a consequence, a DNN outputting two values $\vec{y}_{\text{pred}} = (\mu_{\text{pred}}, \sigma_{\text{pred}})$ and trained with the loss:

$$\mathcal{L}_{\text{MDN}} = -\log(P(\vec{y}_{\text{true}}, \vec{y}_{\text{pred}})) = \log(\sigma_{\text{pred}}) + \frac{1}{2} \frac{(r_E - \mu_{\text{pred}})^2}{\sigma_{\text{pred}}^2} \quad (6.6)$$

would minimise the logarithm of Eq. (6.5) and thus predict the estimators, that is $(\mu_{\text{pred}}, \sigma_{\text{pred}}) = (\hat{\mu}, \hat{\sigma})$. Using this MDN loss, the DNN has to predict the width of this distribution (σ_{pred}) in addition to predicting the mode of the response distribution. For an energy and mass calibration DNN:

$$\vec{y}_{\text{pred}} = \begin{pmatrix} \mu_{\text{pred}}^E \\ \sigma_{\text{pred}}^E \\ \mu_{\text{pred}}^m \\ \sigma_{\text{pred}}^m \end{pmatrix},$$

where only the predicted modes are eventually used in the final calibration procedures.

In practice, the energy or mass response distributions are not perfect Gaussian distributions especially for jets with low p_T and high mass or with high $|\eta|$. To take this into account, the assumed underlying distribution can be replaced by an asymmetric Gaussian distribution or a truncated Gaussian distribution.

$$P_{\text{asym}}(x) \sim \begin{cases} e^{(x-\mu)^2/2\sigma_1^2} & \text{if } x < \mu \\ e^{(x-\mu)^2/2\sigma_2^2} & \text{if } x \geq \mu \end{cases} \quad P_{\text{trunc}}(x) \sim \begin{cases} e^{(x-\mu)^2/2\sigma^2} & \text{if } |x - \mu| < N\sigma \\ 0 & \text{otherwise} \end{cases} \quad (6.7)$$

Although not physically motivated these losses help with the numerical convergence of the network. In particular, the asymmetric MDN loss (MDNA) helps to avoid biases from asymmetric shapes in the response distributions, but it requires the DNN to output three values per prediction (μ, σ_1, σ_2). The use of a truncated distribution can apply to both the MDN and MDNA cases. In practice, truncation consists of excluding from the loss evaluation jets for which $|x - \mu| > N\sigma$ and adapting Eq. (6.6) with the relevant distribution normalisation factor. Asymmetric loss helps to converge faster in regions of phase space where the distributions are asymmetric and the truncation is used at the end of the training to improve the accuracy of the calibration. Details of the choice and tuning of the parameters of the loss function are further discussed in Section 6.4.4.

6.4.2 Input features

To obtain precise predictions, it is necessary to give the DNN an as complete description of the jet and the event as possible. A list of 21 jet or event variables is chosen as the input to the DNN. These features are chosen for their relevance in calibration or jet identification tasks as demonstrated in previous studies [80, 104]. They can be divided in four categories: the jet kinematics, which are the main dependencies of the energy and mass responses³; the jet substructure variables, which describe the particle-shower topology used to classify jets; the detector variables, which quantify the relative charged-particle contribution and the development of the calorimeter showers; and finally, the event variables that describe the jet environment. Table 6.2 presents these input variables and their definitions. To have homogeneous inputs for the DNN, each input is normalised so that its distribution falls in the range $[-1, 1]$ ⁴.

In addition, the DNN has difficulties learning the sharp variations in the energy and mass responses along η that are due to the complex ATLAS detector geometry and in particular to the poorly instrumented regions between the central and forward calorimeters between $|\eta| = 0.8$ and $|\eta| = 1.5$.

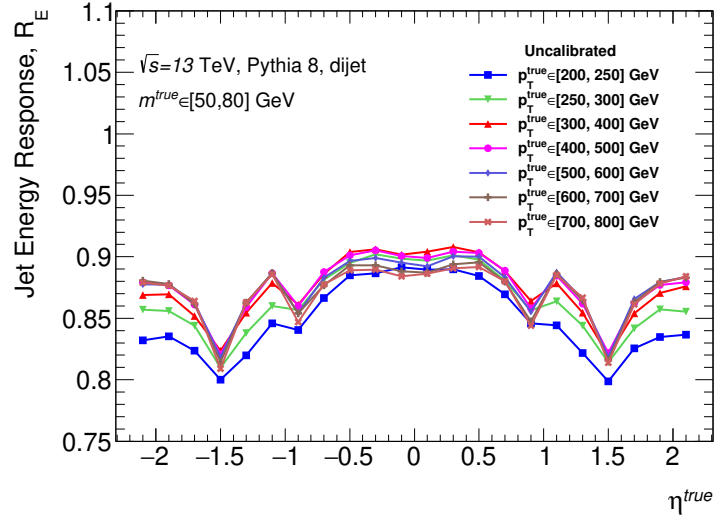
This is seen in Figure 6.4 which shows off the energy response before any calibration as a function of η . The sharp variations seen around $|\eta| = 1$ and $|\eta| = 1.4$ are difficult to model when training a DNN: when close to a transition region, small variations of η produce important response variations whereas away from these regions, the dependency on η is small. These two regimes and the fast transition between them is something to which the DNN has difficulty adapting, having only one information on the jet position inside the detector.

³ Although the rapidity provides a better estimate of the actual angular position of massive objects, the pseudorapidity is used in this paper as used in the current ATLAS calibration; the performance is not expected to be affected by this choice.

⁴ This is done by applying a scale factor s and adding an offset o : $f(x) = s \times x + o$. The jet energy and mass are also transformed before the normalisation by taking their logarithms.

Table 6.2: The input features of the DNN.

	Name	Definition
Jet level	E	Energy of the jet in GeV, $\log E$ is taken to reduce the spread of its distribution
	m	Mass of the jet in GeV, $\log m$ is taken to reduce the spread of its distribution
	η	Jet pseudorapidity
Substructure level	groomMRatio	Mass ratio of groomed to ungroomed jets
	Width	$\sum_i p_{Ti} \Delta R(i, \text{jet}) / (\sum_i p_{Ti})$ where ΔR is the angular distance (sum over the jet constituents)
	Split12, Split23	Splitting scales at the 1st and 2nd exclusive k_T declusterings [105]
	C2, D2	Energy correlation ratios [106, 107]
	τ_{21}, τ_{32}	N-Subjettiness ratios using WTA axis [108, 109]
Detector level	QW	Smallest invariant mass among the proto-jets pairs of the last 3 steps of a k_T reclustering sequence
	EMFrac	Energy fraction deposited in the electromagnetic calorimeter
	EM3Frac	Energy fraction deposited in the third layer of the electromagnetic calorimeter
	Tile0Frac	Energy fraction deposited in the 1st layer of the hadronic calorimeter
	EffNConsts	$(\sum_i E_i)^2 / (\sum_i E_i^2)$ (sum over the jet constituents)
	NeutralFrac	Energy fraction from neutral constituents
	ChargedPTFrac	p_T fraction from charged constituents
Event level	ChargedMFrac	Mass fraction from charged constituents
	μ	Mean number of interactions per bunch crossing
	NPV	Number of primary vertices per event


 Figure 6.4: Uncalibrated Jet Energy Response as a function of η^{true} for jets with $m^{\text{true}} \in [50, 80]$ GeV.

In an effort to help the modelling, an additional input preparation step called η annotation is implemented as the first operation of the DNN. This step computes extra features based on η that are passed along with other input features to the rest of the DNN. These extra features are calculated as $f_i(\eta)$ where f_i are Gaussian functions with different central values and widths that are chosen to cover key angular regions of the detector (such as the central part or edges of the electromagnetic barrel calorimeter and of other calorimeters). Figure 6.5 shows how the 11 different Gaussians functions are positioned, providing a continuous encoding of the proximity of the jet to these regions. These processed values are included alongside the unprocessed η for a total of 12 representations.

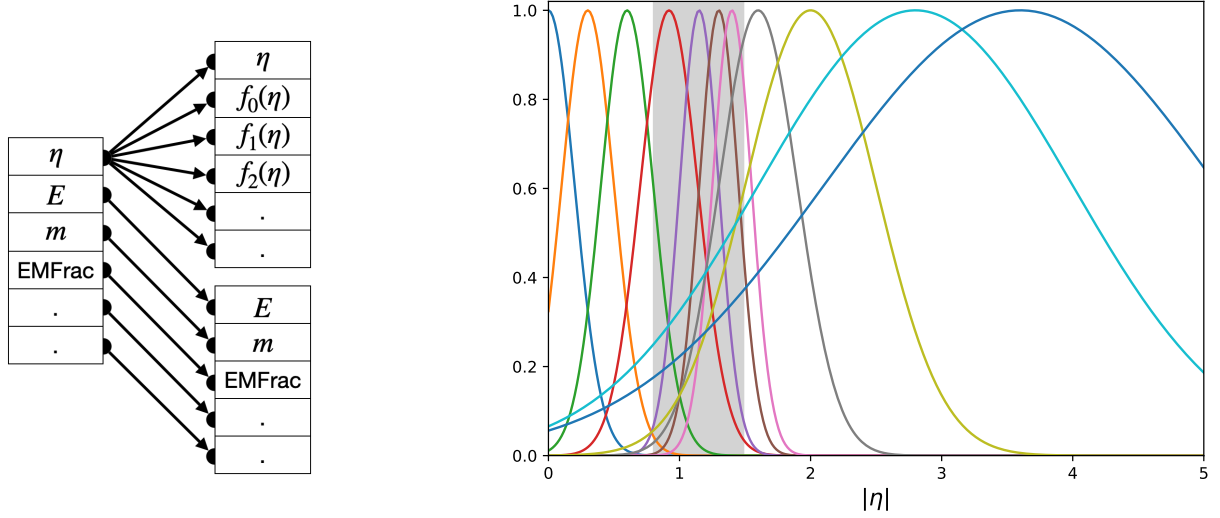


Figure 6.5: The principle of the η annotation (left) and the various $f_i(\eta)$ functions (right). The vertical shaded band on the right plot marks the calorimeter transition region.

An alternative to the η annotation could have been to train different networks in different η bins. But beside the additional technical complexity needed to deal with multiple NN, this would induce discontinuities along the η parameters unless even more complexity is implemented to interpolate between η bins.

6.4.3 Architecture

The DNN architecture is mainly based on two specific parts, the η annotation (described in Section 6.4.2) and a fork between the energy and mass outputs. Both are designed to address the main difficulties of the jet calibration, the η dependency of the responses and the important differences between the energy and mass response distributions. Building on this base, the architecture was iteratively improved and tested to obtain an optimal structure providing good performance. Consequently, the DNN has a relatively complex structure as represented in Figure 6.6, where:

- *Dense* (N) blocks represent the matrix, bias and activation function resulting in a hidden layer with N neurons.
- The *Concatenate* block is a static layer (without trainable parameters) that concatenates its two input layers.

- The *Multiply* block is a static layer that performs an element-wise multiplication of its input layers.
- The final *Output* layers give the predictions for the energy and mass responses. They may consist of two or three neurons depending on the loss function (as discussed in Section 6.4.1).

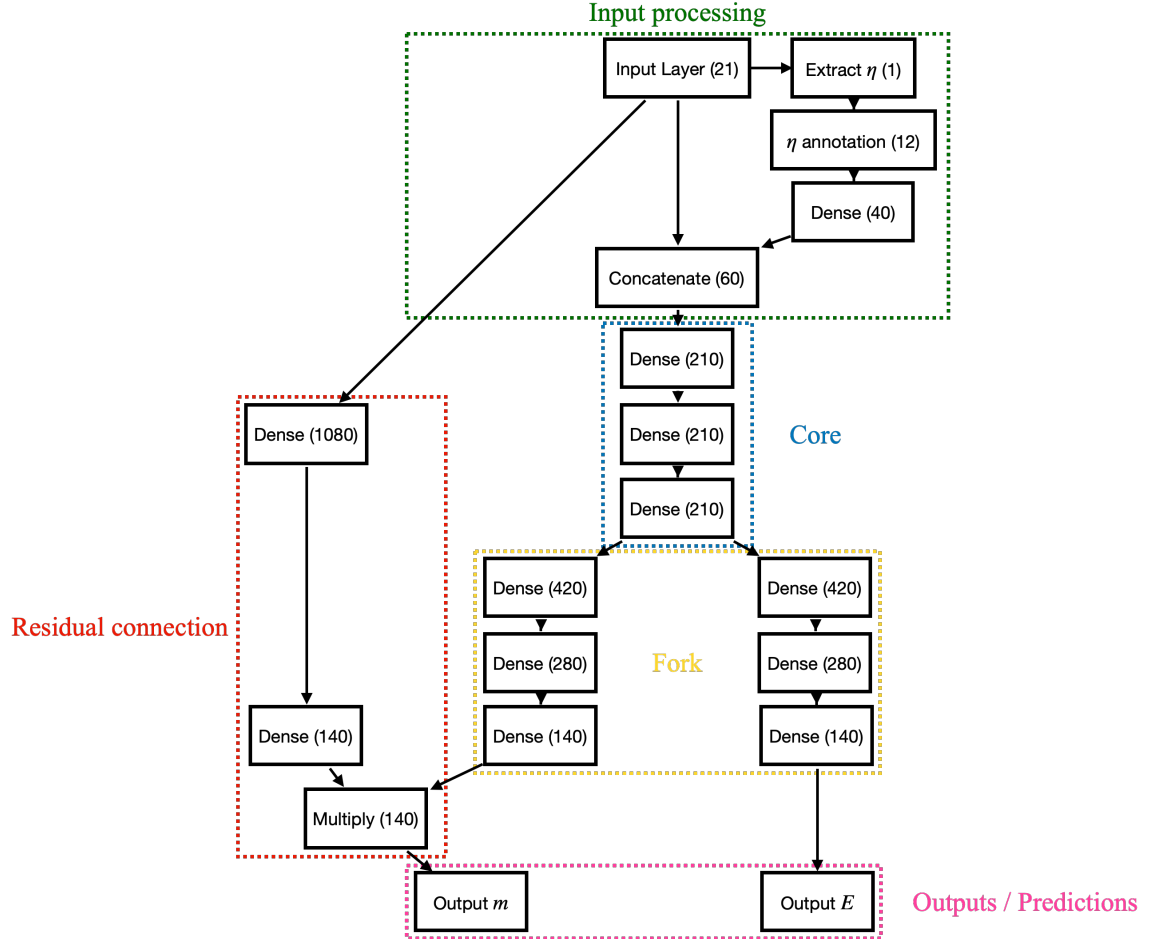


Figure 6.6: The architecture of the DNN.

The DNN contains approximately 764 000 trainable parameters and is structured in different parts as indicated in Figure 6.6.:

- **Input processing:** contains the input layer with the 21 input features and the η annotation described above.
- **Core:** contains several hidden layers (210 neurons each) common to both the energy and mass predictions.
- **Early fork:** premature splitting between the energy and mass predictions to allow for a strong specialisation of the DNN in each calibration, both forks containing the same hidden layers. Having independent weights also allows to *freeze* some of them during the training procedure.
- **Multiplicative residual connection:** loosely inspired from Ref. [110] with the substitution of the addition by a multiplication to increase the class of functions that can be modelled by the network.

These layers link the input layer directly to the mass output, with the intention of making the DNN learn which inputs are the most important for the mass calibration and thus helping to improve the mass predictions.⁵

- Outputs: the DNN has two output predictions, one for the energy calibration and one for the mass calibration, $\vec{y}_{\text{pred}} = (\mu_{\text{pred}}^E, \mu_{\text{pred}}^m) = (R_E^{DNN}, R_m^{DNN})$, along with subsidiary predictions $(\sigma_{\text{pred}}^E, \sigma_{\text{pred}}^m)$.

The activation function for each hidden layer is chosen to be the MISH function [111] and is given by:

$$f(x) = x \tanh(\ln(1 + e^x)).$$

The only exception is the activation function of the last hidden layer of the residual connection that is the *softmax* function: its outputs are restricted to $[0, 1]$ and are used as multiplicative weights.

The last activation function applied to the output layers is chosen to be:

$$f(x) = \tanh(x) + 1$$

to obtain predictions between 0 and 2, which is the expected range for the energy and mass central responses.

6.4.4 Training procedure

Establishing the training strategy cannot be left to a simple hyperparameter scan. Firstly, because the energy and mass responses vary significantly across the phase space that is not uniformly populated and, secondly, because converging to a loss minimum does not guarantee that the physics performance goals are satisfied for the full phase space. Furthermore, these performance goals (described in Section 6.5) are costly to evaluate and cannot be tested during the training procedure. As a consequence the strategy presented here is obtained with a series of attempts and refinements gradually improving the performance. This training procedure is described in Table 6.3: the DNN is trained on the QCD PYTHIA dijet sample, containing around 270 million jets, for a total of 107 epochs divided into three main steps.

In a first initialisation step, the DNN is trained with small batch sizes and an MDNA loss with no or small truncation of the loss. This aims at quickly converging to a rough estimate of the mode of the response distributions thanks to small batch size (i.e., frequent weight updates) and a few epochs while adapting to possible asymmetric distributions (e.g., at high η) with the MDNA loss.

The second step is designed to obtain very precise predictions by taking advantage of large batch sizes and a progressive narrowing of the loss truncation. This narrowing allows a gradual focus on the centre of the response distributions, improving the predictions. The larger batch sizes improve the calculation speed and thus the training time.

The third step is a fine-tuning step with an exclusive mass training. The weights involved in the energy prediction⁶ are frozen to their values at the end of the second step and only those involved in the mass

⁵ This is only added for the mass calibration prediction as it is harder to predict than the energy calibration due to its response distributions being broader and more asymmetric. The predicted energy calibration already shows excellent performance without an additional residual connection.

⁶ These comprise the weights involved in the input processing, the core and the energy-dedicated fork of the network architecture.

prediction⁷ are free to be tuned. This step is necessary in order to have as good predictions for the mass as for the energy for which the two first steps are enough. For this last step, a MDN loss with a large truncation at 1.0σ is used in order to focus more on the core of the mass response distribution. A symmetric loss at this stage is observed to converge better, possibly because it results in a simpler network, easier to optimize on the central and more symmetric parts of the response distributions.

During the training two different optimizers are used, *Rectified Adam* [112] in the initialisation step and *diffGrad* [113] in the common training and exclusive mass training. The learning rate is decreased in the latter steps of the training following the improvements in the predictions.

Table 6.3: The specific training procedure of the DNN. MDN(A) stands for mixture density network (asymmetric). The loss truncation is indicated in number of standard deviations, σ . When E and m are not mentioned, it means that the energy and mass losses are identical.

Steps	N°	Number of epochs	Batch size	Learning rate	Loss
Initialisation	1	2	15 000	10^{-3}	MDNA
	2	2	25 000	10^{-3}	MDNA
	3	2	35 000	10^{-3}	MDNA truncated (4.0σ)
	4	2	15 000	10^{-3}	MDNA truncated (3.5σ)
Common training	5	6	95000	10^{-3}	MDNA truncated (3.5σ)
	6	6	95 000	10^{-3}	MDNA truncated (3.5σ)
	7	6	125 000	10^{-3}	MDNA truncated (3.2σ)
	8	6	125 000	10^{-3}	MDNA truncated (3.2σ)
	9	10	155 000	$5 \cdot 10^{-4}$	MDNA truncated (3.0σ)
	10	15	95 000	10^{-5}	MDNA truncated ($E: 3.0\sigma, m: 2.0\sigma$)
Exclusive mass training	11	50	95 000	10^{-5}	MDN truncated (1.0σ)

6.4.5 Implementation details

The implementation of the DNN and the training procedure is written in python using the TensorFlow and Keras frameworks [102]. The training dataset is read from the ROOT file format using the *uproot* library [114] All the code can be found in the CERN *GitLab* repository [115] along with some documentation. Trainings were done at the IN2P3 computation centre of Lyon, CCIN2P3, on Nvidia V100 GPUs.

⁷ These comprise the weights in the mass-dedicated fork and in the residual connection layers of the network architecture.

6.5 Performances and validation of the calibration

The evaluation of the DNN performances and the validation of the predicted energy and mass calibrations are presented in this section. Unless mentioned otherwise, these are evaluated on the dijet sample, hence for standard quark and gluon jets. First, some technical performances are inspected in Section 6.5.1, then an overview of the DNN calibration performances is given in Section 6.5.2. Finally, a validation of the DNN calibration is presented in Section 6.5.3.

6.5.1 Technical performances

6.5.1.1 Overfitting

If trained over too many epochs, the DNN can start to learn the particular specificities of the training dataset and then have difficulties to generalize the predictions to other datasets. This phenomenon, called overfitting, must be avoided as the goal is to calibrate any large- R jet. In order to check for possible overfitting, the loss function of the DNN is measured after each epoch on a validation dataset and compared against the training loss (evaluated on the training dataset). This validation dataset is composed of jets which the DNN will never train on. At each epoch end, the validation dataset is fed to the DNN and the losses computed and averaged on it. As long as the loss obtained from the validation dataset stays close to the loss computed on the training dataset⁸, the DNN is considered to generalize well to new input jets and to not overfit the training dataset.

For this study, the validation dataset is a subset of the training dataset that is set aside at the beginning of the training (about ten million jets corresponding to $\sim 4\%$ of the training dataset). Figure 6.7 shows the evolution of the loss with the number of epochs (i.e, the number of jets processed). The training and validation losses (obtained for the energy, the mass and their total) are shown as solid and dotted lines, respectively. The different losses display very good agreement, showing that the DNN does not overfit the training dataset. This was expected as the training dataset is enormous: it would take many more epochs to overfit this dataset. It also shows that the loss does not evolve a lot after two third of the training. For a more common DNN setup it could be wise to stop the training earlier as it seems, by that metric alone, that its performance is not improving any more. However, it has been seen that for this particular problem the predicted calibrations continue to improve even if the loss does not. This was one of the important difficulties for finding a good training procedure (as mentioned in the previous section).

6.5.1.2 Convergence

As a DNN is always initialized with a random set of parameters $\vec{\theta}$, DNNs with the same architecture, loss and training procedure will be unique in terms of parameters. They should nevertheless provide similar performances: this convergence must be verified. This was done by training multiple independent DNNs and checking that the results of the extensive calibration performance validation remain almost identical.

⁸ The training loss is computed at the end of each batch.

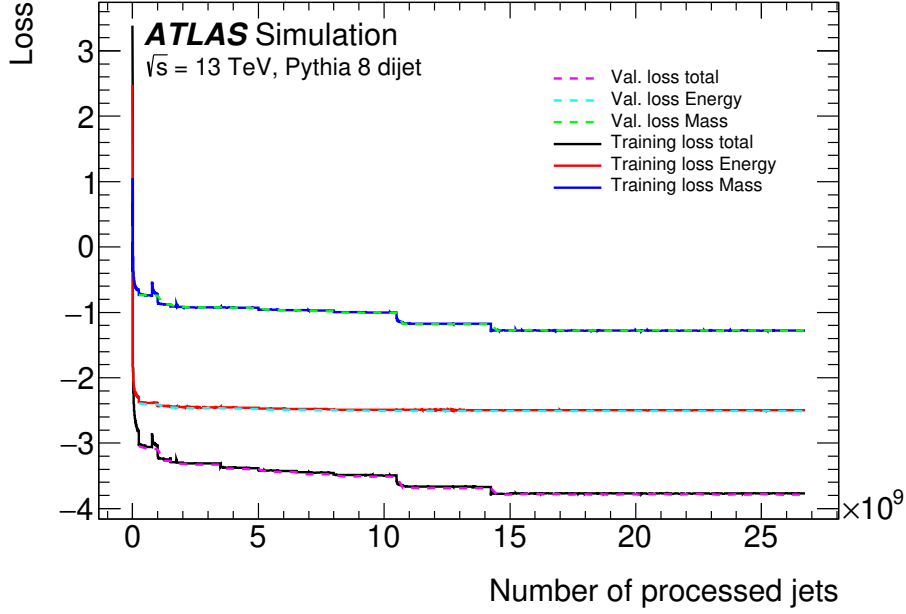


Figure 6.7: The loss evolution for the training (plain lines) and validation (dotted lines) data samples. The different spikes and drops correspond to the transitions between training steps (see Table 6.3) and are due to the variations of the loss function.

6.5.2 Calibration performances

In order to quantify the performances in terms of energy and mass calibrations, ATLAS uses two main measures: the response closure and the resolution of the calibrated response. These two quantities are described in the next section and are then used to compare the DNN calibration performances to the ones obtained with the ATLAS standard calibration. The calibration performances are measured on both QCD jets which were used for training and on others jet topologies (boosted W/Z, Higgs and top jets). The goal is to obtain performances which are better than or are at least similar to those of the standard calibration for large- R jets in the usual phase space in which they are used in physics analysis: $p_T > 200$ GeV, $m > 40$ GeV and $|\eta| < 2$ ⁹.

6.5.2.1 Response closure and resolution

From the predicted energy and mass response $y_{\text{pred}} = (R_E^{\text{DNN}}, R_m^{\text{DNN}})$, the individual energy (or mass) calibrated response is defined as :

$$r_{E_{\text{calib}}} = \frac{E_{\text{calib}}}{E_{\text{true}}} = \frac{1}{R_E^{\text{DNN}}} \frac{E_{\text{reco}}}{E_{\text{true}}} = \frac{r_E}{R_E^{\text{DNN}}}$$

⁹ These kinematic requirements are set by the trigger limitations (low- p_T objects can't be triggered on without the expense of excessive high rates), the coverage of the inner detector (full coverage up to $\eta = 2$) and the purpose of the large- R jets (reconstruction of massive boosted particle decays).

Following the discussion in Sec. 6.1, $r_{E_{\text{calib}}}$ values are realizations of a distribution P from which the calibrated response can be obtained : $R_{E_{\text{calib}}}(\vec{x}_{\text{true}}) = \text{mode}(P(r_{E_{\text{calib}}} | \vec{x}_{\text{true}}))$. To evaluate $R_{E_{\text{calib}}}(\vec{x}_{\text{true}})$ the individual responses $r_{E_{\text{calib}}}$ are collected in bins of $\vec{x}_{\text{true}}$ (where typically $\vec{x}_{\text{true}} = (p_T^{\text{true}}, m^{\text{true}}, \eta^{\text{true}})$ but other binning can be chosen depending on the test). In these bins the individual calibrated responses are expected to be gaussian distributed, hence a gaussian fit is performed to obtain the mode and the width of the calibrated response as shown in Figure 6.8. Technically, a smoothing of the histogram is applied with a Savitzky-Golay filter [116] before the fit to improve its stability in low-statistics bins. Heuristics are also employed to limit the fit around the central value (by setting the fitting range around the mean of the histogram between $\pm N\sigma_{\text{hist}}$, with σ_{hist} the standard deviation of the histogram).

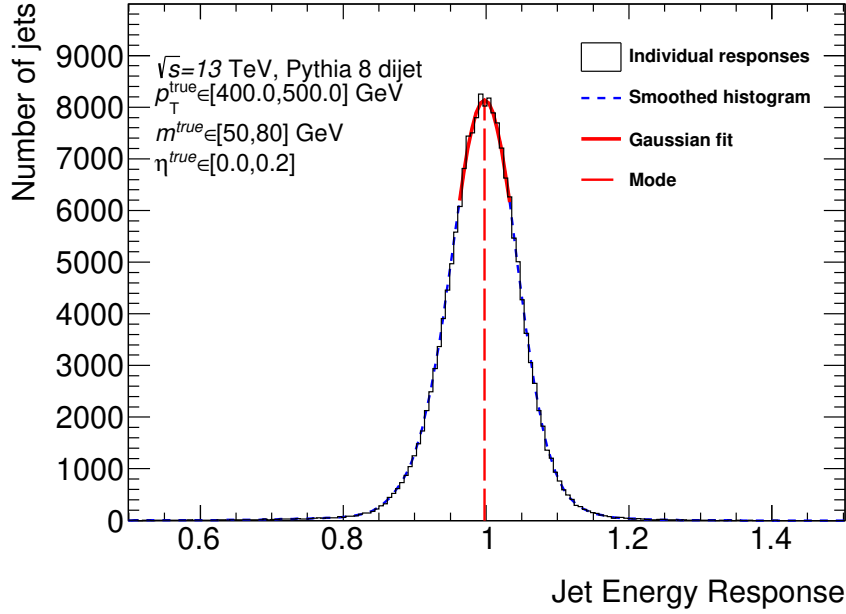


Figure 6.8: Fit of the calibrated energy response for jets with $400 < p_T^{\text{true}} < 500$ GeV, $50 < m^{\text{true}} < 80$ GeV and $0.0 < \eta^{\text{true}} < 0.2$.

This procedure allows to evaluate the relevant quantities in each bin of $\vec{x}_{\text{true}}$:

- the calibrated response $R_{E_{\text{calib}}}$ or $R_{m_{\text{calib}}}$ is the central value of the gaussian fit; it should be as close to one as possible. The response closure is considered to be performant when the calibrated energy (mass) response is within $\pm 1\%$ ($\pm 5\%$) of unity, for all bins.
- the calibrated response resolution σ is taken as the response distribution's 68% IQR divided by the calibrated response as defined in Eq. 6.1; it should be as small as possible.

The performances are then visualized as graph of these quantities as a function of a variable defining the bins (for example as a function of the jet p_T for a specific bin in mass and η).

6.5.2.2 Effect of η annotation

Section 6.4.2 describes the necessity to add the η annotation to help the DNN adapt to the complex detector geometry. The assumption is that the extra information represented by the angular encoding together with the related weights allows the training process to specialise the predictions according to the detector region. As a validation step, Figure 6.9 shows a comparison of the jet energy and mass response closure between two DNN setups with the same training procedure and the same architecture except for the η annotation. Its inclusion clearly improves the DNN calibrated response closure in the almost complete η range and especially in the calorimeter transition region above $|\eta| = 1.2$. The JMS improvement is less important, as expected since the uncalibrated response (squared markers) variations for the mass are less important than for the energy.

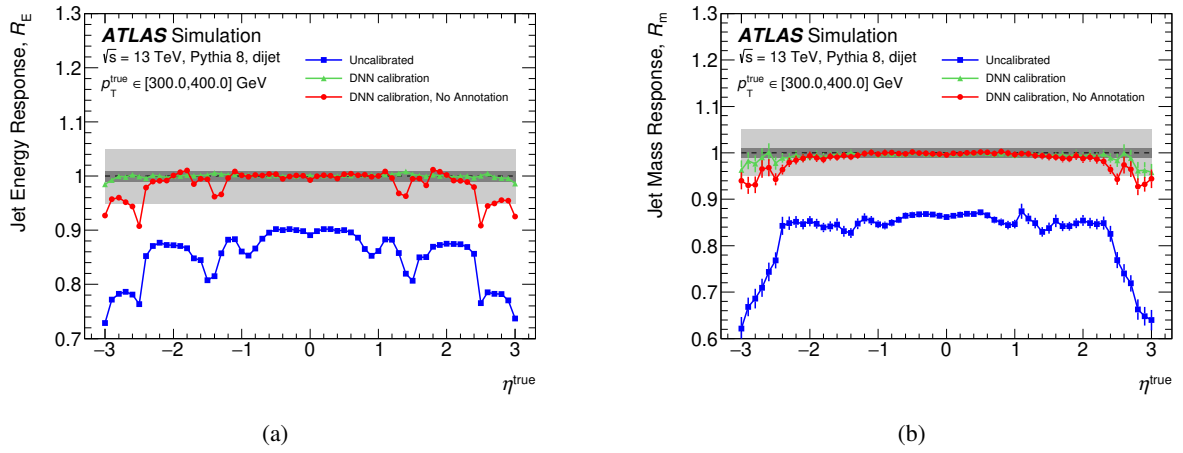


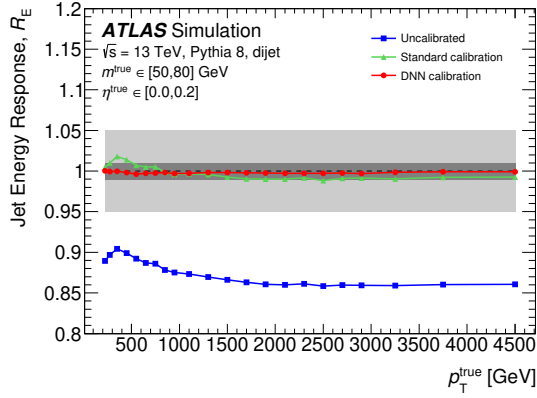
Figure 6.9: The jet energy (a) and mass (b) response before calibration (squares) and after the DNN calibration with (triangles) or without (circles) the η annotation step as a function of η^{true} . The horizontal bands represent 1% (dark) and 5% (light) up and down divergences from closure.

These improvements are present in every p_T and η bin, proving the importance of the η annotation step.

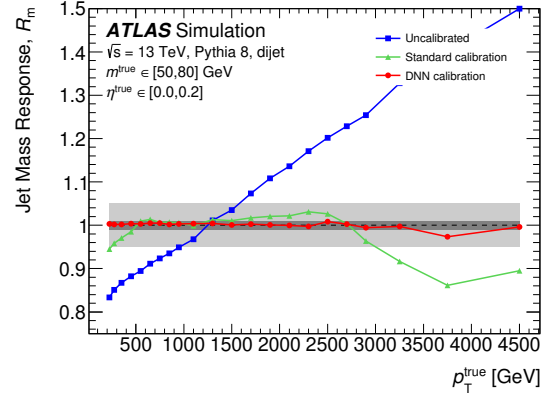
6.5.2.3 Performances on QCD dijets

As a closure test, the DNN calibration is applied to the same QCD sample used during training. This allows verification that the DNN convergence translates into good physics performance in a sample with large statistics. Overall, the calibration is found to perform very well for $|\eta| < 2$ with the best results at high mass in terms of response closure and resolution. Figure 6.10 presents the energy and mass calibrated responses as a function of p_T^{true} and η^{true} and the energy and mass resolution as a function of p_T^{true} for specific bins. It shows excellent response closure for both energy and mass, closer to unity than the standard calibration [80]. The DNN calibration also shows an important improvement in the resolution relative to the standard calibration. For the mass, the uncalibrated resolution appears better than the calibrated ones. However, this is a relative quantity (as defined in Section 6.1) while the absolute value of the IQR scales with the central value. As a consequence, comparing the uncalibrated resolution corresponding to responses significantly differing from 1.0 with a calibrated resolution is not very meaningful.

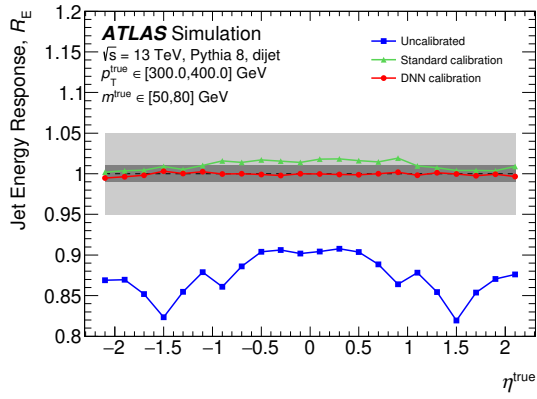
Figure 6.11 presents the excellent consistency of the DNN calibration performance for different p_T^{true} bins.



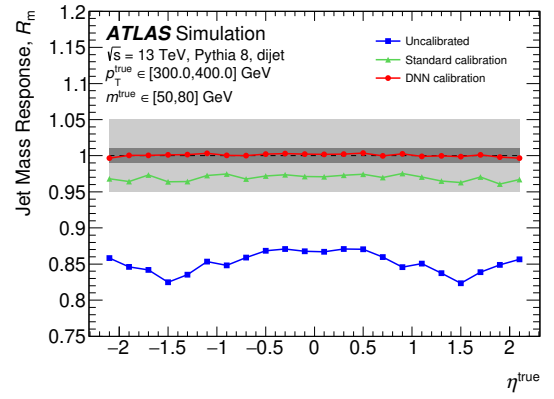
(a)



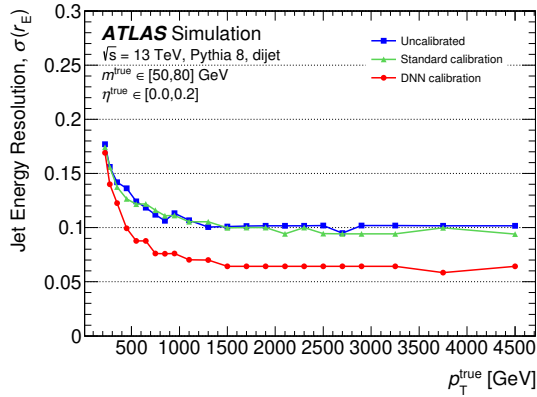
(b)



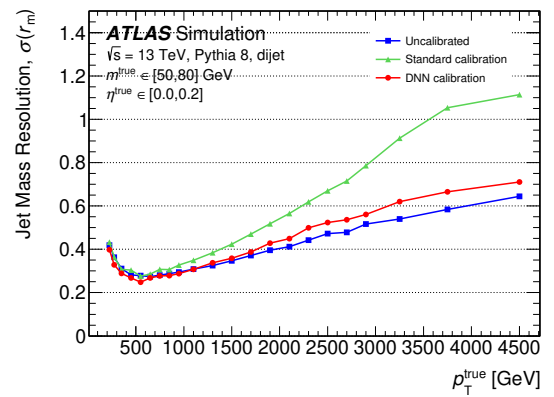
(c)



(d)



(e)



(f)

Figure 6.10: Jet energy (left) and mass (right) responses (a, b, c, d) and resolutions (e, f) for different calibrations (squares: no calibration, triangles: standard calibration, circles: DNN calibration) as a function of p_T^{true} (a, b, e, f) and η^{true} (c, d). The horizontal bands in (a, b, c, d) represent 1% (dark) and 5% (light) up and down divergences from closure.

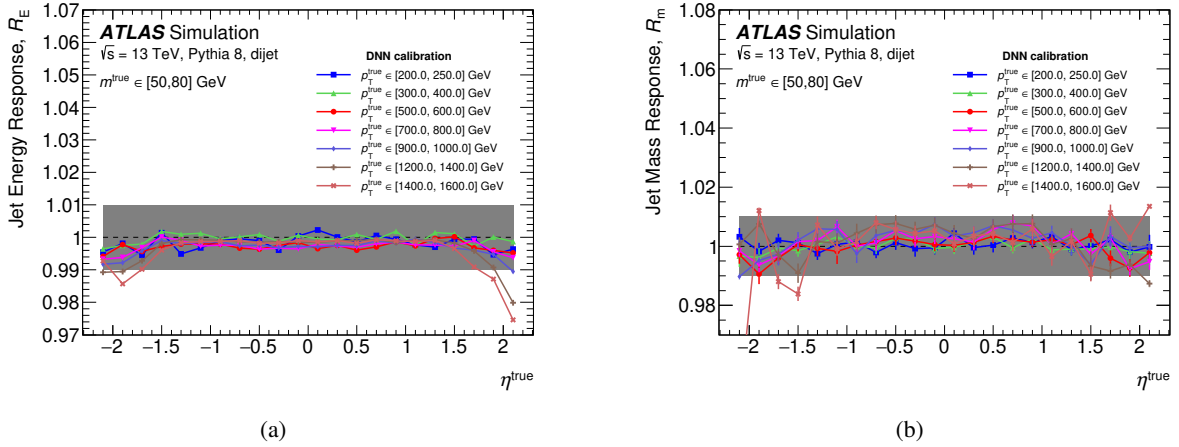


Figure 6.11: The (a) jet energy and (b) mass calibrated response for the DNN calibration as a function of η^{true} and for multiple bins in p_T^{true} . The horizontal band represents 1% up and down divergence from closure.

In every bin of $(p_T^{\text{true}}, m^{\text{true}}, \eta^{\text{true}})$ with $p_T^{\text{true}} > 200$ GeV, $m^{\text{true}} > 40$ GeV and $|\eta^{\text{true}}| < 2$, the DNN calibration is equally or more performant than the standard calibration with a better response closure and resolution. Improvements of a few percent up to around 10% are observed in both energy and mass response closure and resolution.

6.5.2.4 Performance on other topologies

Physics analyses mainly use large- R jets to reconstruct the hadronic decays of massive and boosted particles like top quarks, W , Z or Higgs bosons. These decays result in jets with different shower topologies than those originating from the dijet samples with which the DNN is trained; this can have an impact on the mass response as demonstrated in Ref. [87] where significant differences between W -initiated and q/g -initiated jets were observed. It is hence very important to verify that the DNN predictions generalise to other such topologies as the same calibration is applied to all jets in data as their origin will be unknown. For this, simulated samples of processes with heavy hadronically decaying particles, as described in Section 6.2.1, are considered. For each sample, the same jet selection and DNN calibration as for QCD jets are applied and the resulting scale and resolution are evaluated in mass bins centred on the mass of the corresponding massive particle. The population of these massive, high-momentum jets is then expected to be entirely dominated by the hadronic decays of the heavy particles considered so that the effect of the topology can be assessed.

Figure 6.12 presents the jet mass response distributions obtained with boosted massive jets (W/Z , Higgs boson and top-quark decay) for different calibrations (uncalibrated, standard calibration and DNN calibration). It shows the benefits of the DNN calibration compared with the standard calibration or the absence of a calibration; the distributions are narrower with less asymmetric tails and are better centred on 1.0.

Figure 6.13 shows the jet energy and mass average responses as a function of the true p_T for the same samples and calibrations. As expected, large differences between the different topologies are observed before calibration, especially in the mass response. While both the standard and the DNN calibrations reduce these response differences, the DNN calibration outperforms the standard one as it gives better

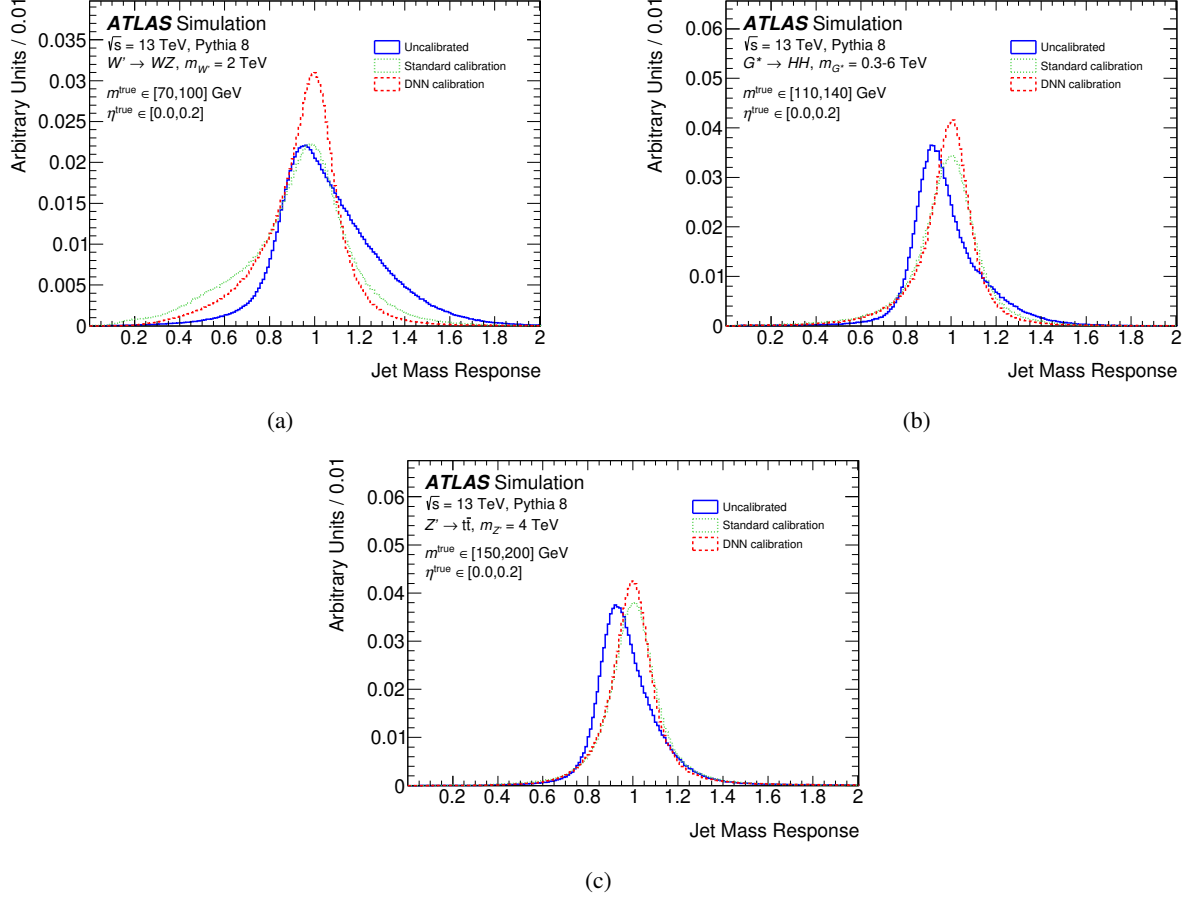
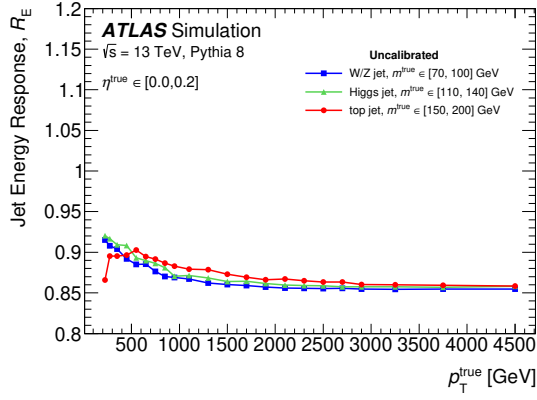
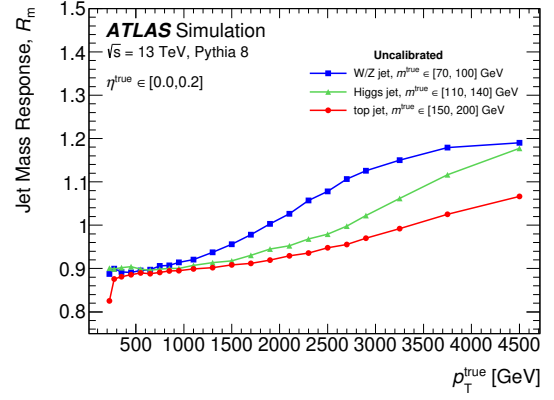


Figure 6.12: (a) Boosted W/Z , (b) Higgs boson and (c) top-quark mass response distributions (normalised to unity) for different calibrations: uncalibrated (plain line), standard calibration (dotted line) and DNN calibration (dashed line).

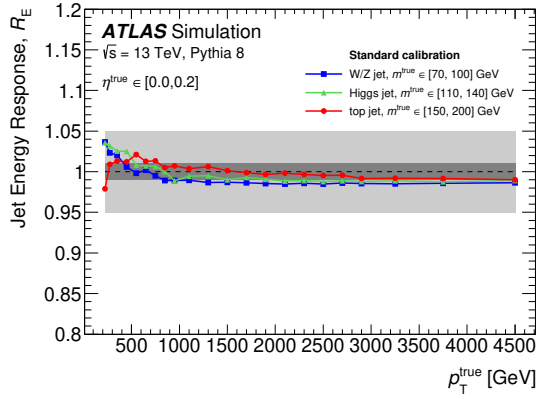
closure for each topology, showing that the DNN predictions generalise well to more complex topologies. A small non-closure in the energy response is however observed for all the boosted topologies at low p_T^{true} (< 500 GeV) for the DNN calibration. For the W/Z and Higgs boson jets, this small under-performance can easily be explained by the internal structure of the boosted massive jets; at high momentum, W/Z and Higgs boson large- R jets are usually composed of two sub-jets with an angular separation of $\Delta R \sim \frac{2m}{p_T}$. At low p_T , their angular separation becomes comparable to the large- R jet radius of 1.0, and the structure becomes neither 1-prong nor 2-prong. Such a structure is possibly rare in the dijet training sample or not captured by the input substructure variables, making it impossible for the DNN to learn to correct for this effect. For top-quark jets, a similar effect could be at play when the 3-prong decay ($t \rightarrow Wb \rightarrow qq'b$) is not fully contained within the jet radius, thus leading to atypical jets. To solve this difficulty, a dedicated additional network structure and training can be envisaged and is discussed in Section 6.6. In addition, for both the standard and DNN calibrations a drop in the mass response can be observed at very high p_T for the W/Z jets. This is due to over-boosted jet systems which lead to a very strong collimation of the jets constituents and ultimately a complete loss of the internal structure of these jets. For such instances, the calibration (standard or DNN) is unable to recover the mass of the W or Z boson as it is experimentally



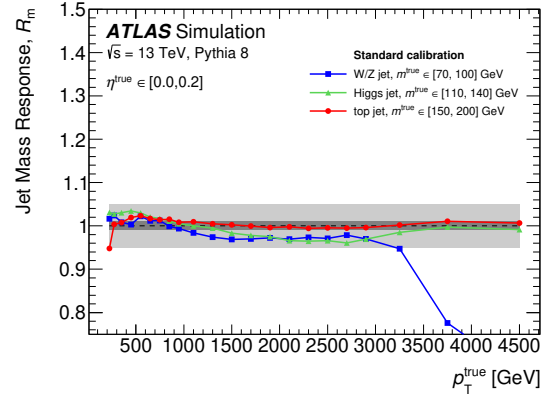
(a)



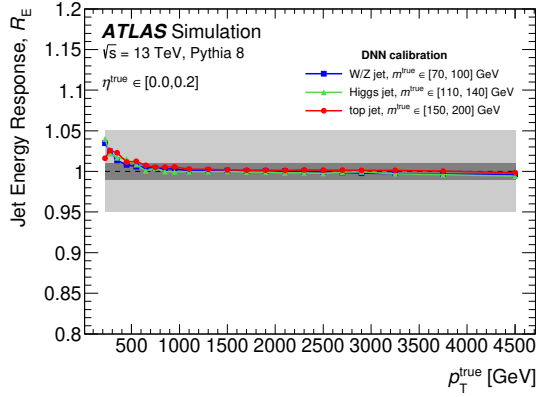
(b)



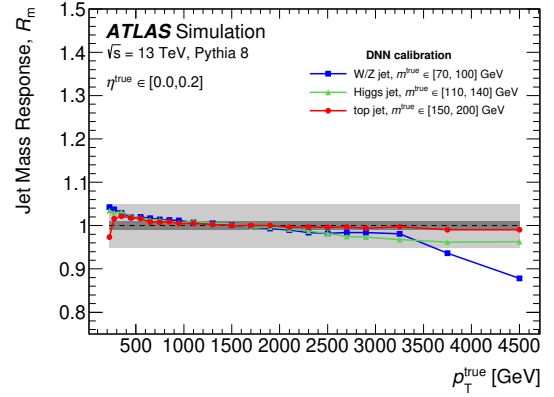
(c)



(d)



(e)



(f)

Figure 6.13: Boosted W/Z, top-quark and Higgs boson jet energy (left) and mass (right) responses as a function of p_T^{true} for different calibrations: (a, b) no calibration, (c, d) standard calibration and (e, f) DNN calibration. The horizontal bands represent 1% (dark) and 5% (light) up and down divergences from closure.

not accessible any more¹⁰. This effect could be corrected by using a calibration specifically derived from boosted W/Z jets and applied only to W/Z jets, however it would, first, not be relevant for data where the origin of the jets is a priori unknown and, second, stand against the idea of a generic calibration.

Figure 6.14 presents the jet energy and mass resolution for the boosted topologies. Here as well, the DNN calibration shows improved performance relative to the standard calibration. The resolution differences between the topologies here can be explained by the increasing mass bins in which the responses are evaluated; the resolution usually gets better at higher masses.

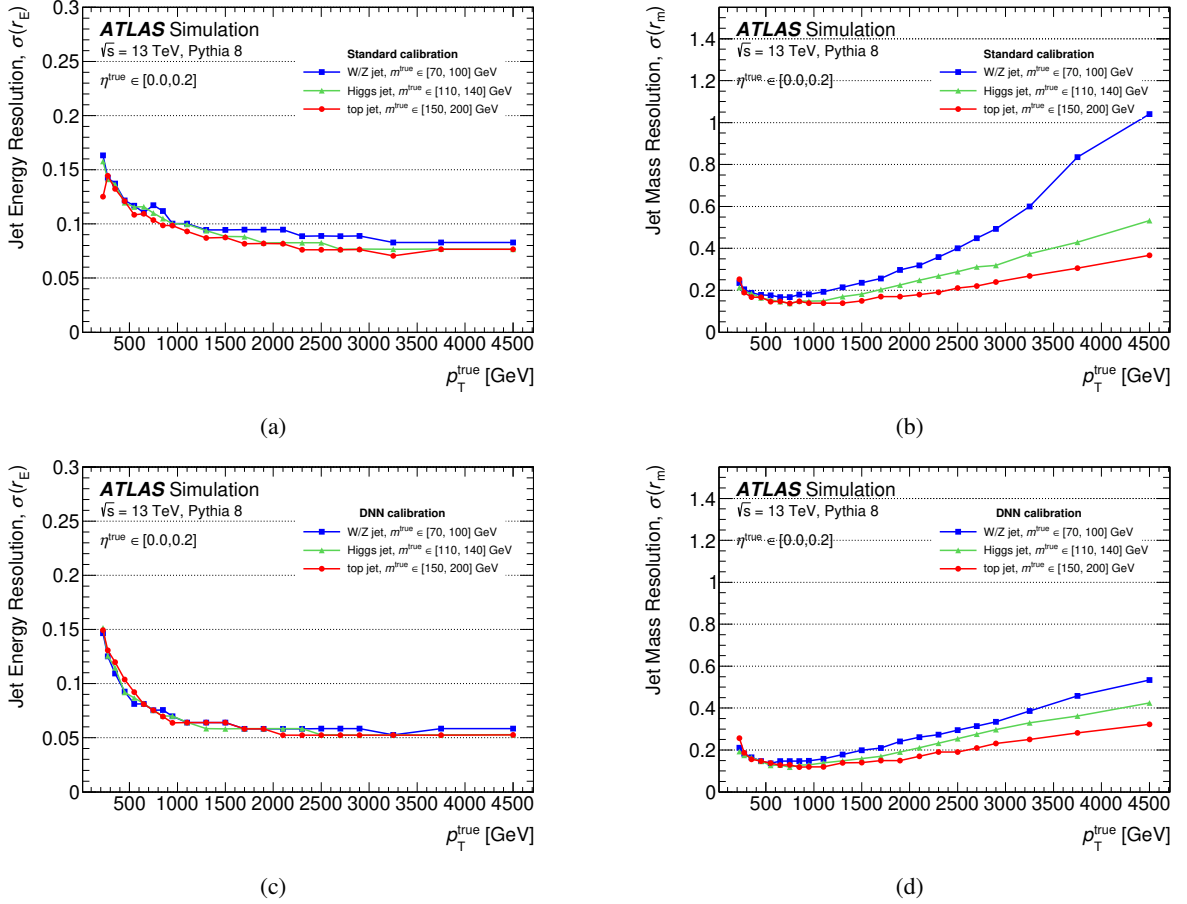


Figure 6.14: Boosted W/Z , top-quark and Higgs boson jet energy (left) and mass (right) resolutions as a function of p_T^{true} for different calibrations: (a, b) standard calibration and (c, d) DNN calibration.

6.5.3 Validation of the DNN calibration

An extensive validation procedure was put in place to ensure the good physical behaviour of the DNN calibration. As above, the same tests were performed with both the ATLAS standard and DNN calibrations and compared. These tests aim at evaluating the calibration dependency on various conditions affecting the jet reconstruction (pileup, underlying input variable distributions, jet isolation criteria).

¹⁰ This effect should also appear for the Higgs boson and top-quark jets, only at higher energies.

Spectrum dependency: The DNN is trained and validated on a specific set of jets produced through a particular Monte Carlo process with a large range of energy. This set of jets leads to specific distributions of the input variables. The DNN calibration performance should be independent of the distribution to which it is applied, so it can be used in different physics analyses with different event selections resulting in different input distributions. Since the calibration is already assessed as a function of true p_T , the impact of the distribution of this variable (or the energy one) is not expected to be strong. However, other input variables are correlated with the p_T and are not binned in the performance tests; their distribution shapes may thus impact the calibration. To verify this, the performance is re-evaluated by changing the response distributions using statistical weights when constructing the response histograms. The weights are chosen to obtain a flat jet energy distribution, entirely different from the original one used in the training as illustrated in Figure 6.15(a)¹¹.

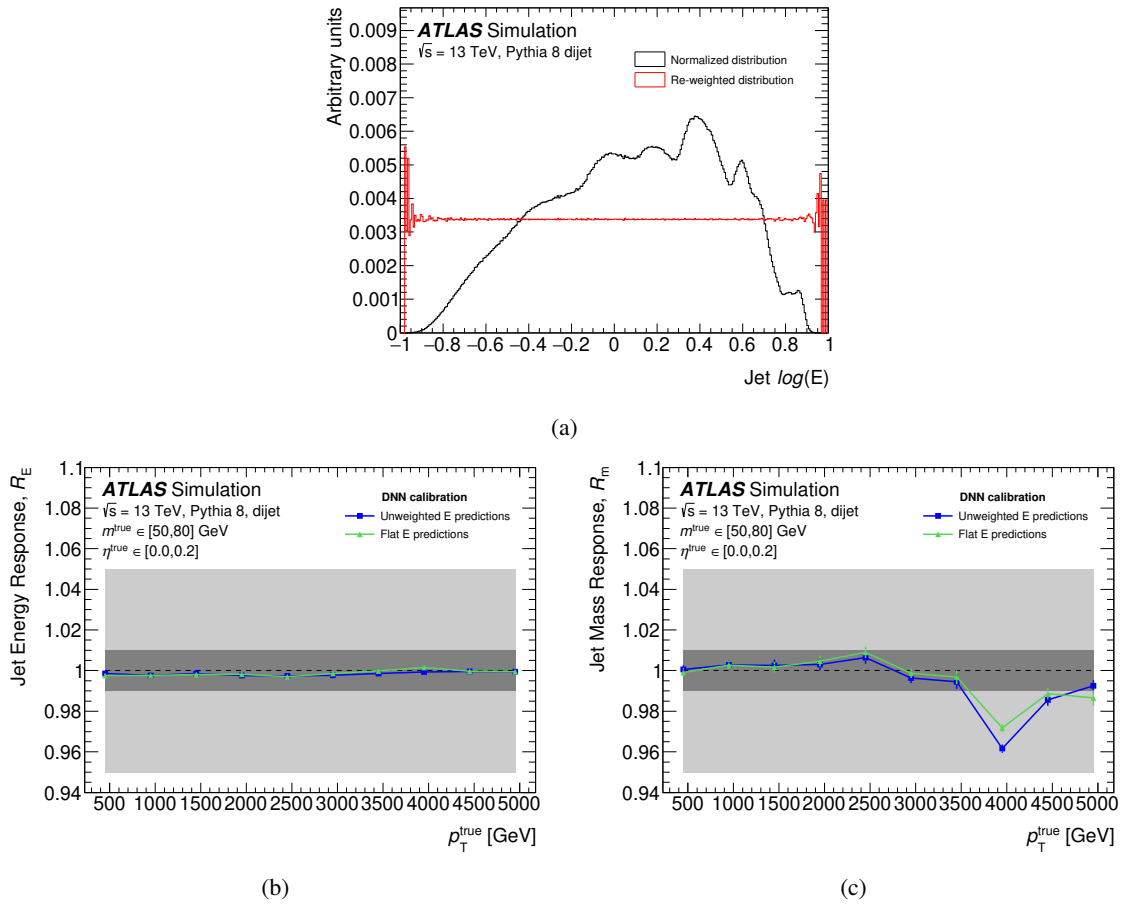


Figure 6.15: Normalised distributions of (a) the jet energy before (black) and after (red) re-weighting. And the corresponding (b) jet energy and (c) mass responses obtained with the DNN calibration as a function of p_T^{true} for these two energy distributions, the unweighted distribution (squares) (as used for training) and the flat-energy weighted distribution (triangles). The horizontal bands represent 1% (dark) and 5% (light) up and down divergences from closure.

The performance is then evaluated in large p_T bins similar to those that physics analysis may use and inside

¹¹ Distributions are shown after transformation and normalisation, e.g., the histogram shows $f(E) = s \times \log(E) + o$ where s and o are such that $f(E)$ is mostly distributed in $[-1, 1]$

which distribution differences may have significant effects. Figure 6.15 (b, c) shows the jet energy and mass responses obtained with the DNN calibration with (square markers) and without (triangular markers) re-weighting. The differences between the two calibrated response closures are very small, proving that the DNN calibration does not depend significantly on the input jet distribution.

Pileup dependency: The impact of the pileup conditions is tested and evaluated by considering the dependency of the response on two quantifying variables: the number of reconstructed primary vertices in an event, N_{PV} , and the average number of collisions per bunch crossing, μ . For this, the energy (or mass) response $R_{E_{\text{calib}}}$ is estimated in bins of $(v_{PU}, E^{\text{true}}, \eta^{\text{true}})$ where v_{PU} is N_{PV} or μ . A linear fit to $R_{E_{\text{calib}}}$ against v_{PU} is then performed for each $(E^{\text{true}}, \eta^{\text{true}})$ and the corresponding slope, the *gradient* relative to v_{PU} denoted as $\frac{\partial R_E}{\partial v_{PU}}$, is extracted. Figure 6.16 presents the gradient relative to N_{PV} of the energy and mass calibrated response modes as a function of η . For both the energy and the mass, the DNN calibration shows smaller gradients than the standard calibration. Similar results are obtained for μ .

This shows that the grooming procedure performed during the jet reconstruction is not enough to suppress pileup effects, while the DNN is able to do so by including N_{PV} and μ as input variables.

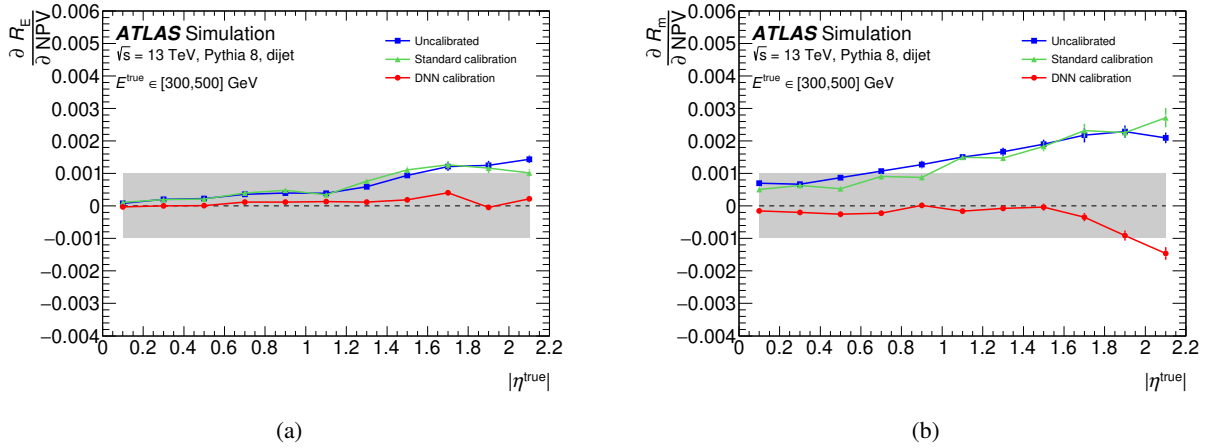
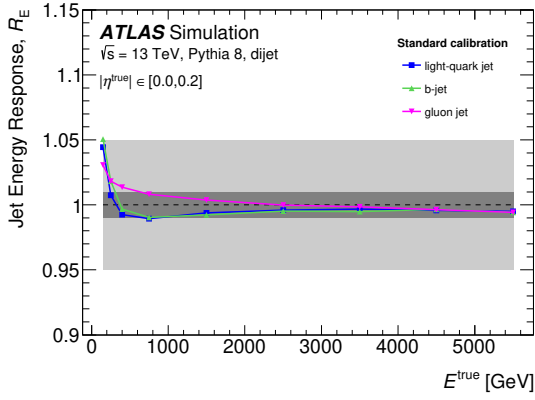


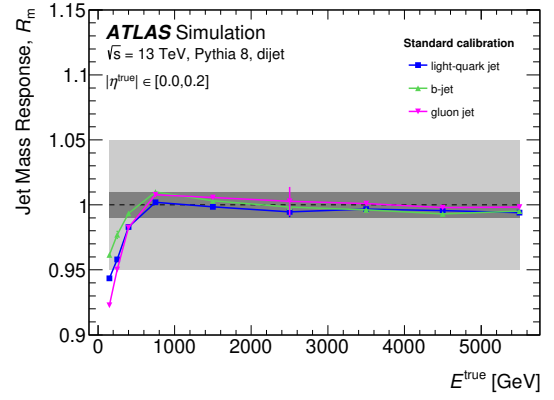
Figure 6.16: The dependence of (a) the jet energy and (b) the mass response on N_{PV} for different calibrations (squares: no calibration, triangles: standard calibration, circles: DNN predicted calibration) as a function of $|\eta^{\text{true}}|$. The horizontal band represent 0.1% up and down divergences from a null gradient.

Jet flavour dependence: As discussed above, the DNN performance is assessed with different topologies (QCD, W/Z , top-quark, or Higgs boson jets). It is also important to understand the dependency of the DNN calibrations on the actual parton that initiated the jet, also known as its flavour (light-quark (u, d, s, c), b -quark or gluon jets), as different responses for different flavours will translate into systematic uncertainties through the uncertainty in the flavour composition of the jets.

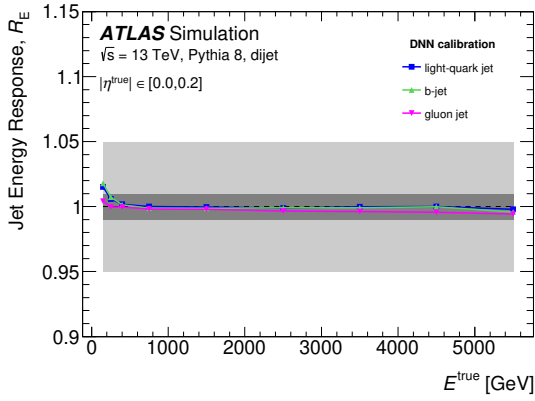
Figure 6.17 presents the energy and mass calibrated responses as a function of the jet energy for different jet flavours, comparing the DNN calibration to the standard calibration. The jet flavour is defined here as the flavour of the parton with the highest energy in the parton shower. These show that the DNN predicts better calibrations for the different partons, with smaller inter-flavour differences relative to the standard calibration, especially at low p_T .



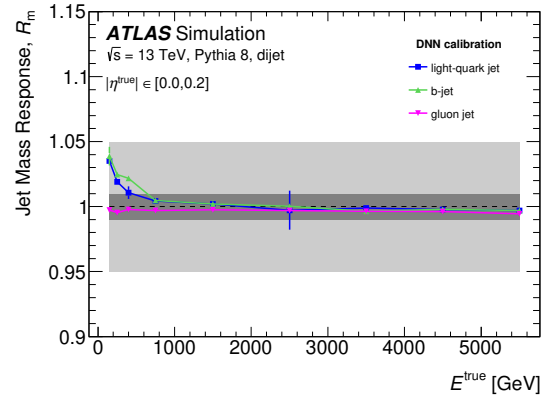
(a) Standard calibration



(b) Standard calibration



(c) DNN calibration



(d) DNN calibration

Figure 6.17: Jet energy (left) and mass (right) responses versus E^{true} for different jet flavours, calibrated with the standard calibration (a, b) and the DNN calibration (c, d). The horizontal bands represent 1% (dark) and 5% (light) up and down divergences from closure.

Jet-isolation dependency: The measurement (or determination) of the jet energy and mass responses can be affected by the angular proximity of neighbouring jets. Such a potential effect on the DNN calibration is tested by predicting energy and mass calibrations with the DNN on a jet sample that is identical to the nominal one (see Section 6.2.2) but with both truth and reconstructed isolation criteria ΔR relaxed from 2.5 to 0.75 and 1.5 to 0.75, respectively. These new criteria lead to low constraints on the jets and thus to a near perfect selection efficiency. Figure 6.18 shows that this alternate jet sample gives identical performances for the responses. Similar identical behaviour is observed for the resolution.

Calibration differences with respect to the ATLAS standard calibration: In order to verify the DNN calibration does not behave oddly on some population of jets, its calibration factors are compared to those given by the ATLAS standard calibration. Two-dimensional histograms of the calibration factors predicted by the DNN ($r_{\text{calib}}^{\text{DNN}}$) and the ones given by the ATLAS standard calibration ($r_{\text{calib}}^{\text{ATLAS, std}}$) are constructed in the binned phase-space ($p_{\text{T}}^{\text{true}}, m^{\text{true}}, \eta^{\text{true}}$). Figure 6.19 presents such histograms in one set of bins. For the energy, as can be expected, the ATLAS standard calibration shows almost constant calibration factors

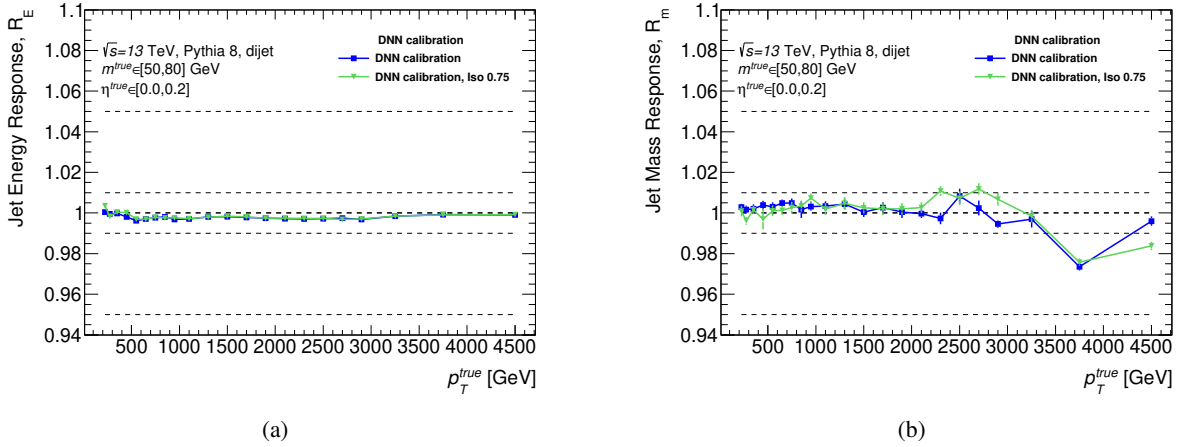


Figure 6.18: Jet energy (a) and mass (b) responses versus p_T^{true} calibrated by the DNN for two different datasets which differ by their jet isolation criterion: the nominal 1.5 (blue) or an alternative 0.75 (green).

in bins of $(p_T^{\text{true}}, m^{\text{true}}, \eta^{\text{true}})$ ¹². While the DNN energy calibration factors spread over a larger range of values showing the flexibility of this new calibration, the average value is close to the ATLAS standard calibration factors and no outlier population is observed. For the mass, it is not as straight forward as the ATLAS standard mass calibration is derived in bins of $\log(m/E)$ and not in bins of the jet's mass only. The ATLAS standard mass calibration factors are therefore not as constant as for the energy. However, the observed mass calibration factors distributions are still centred around the same value and no significant outliers are observed beyond small tails in the distribution. Overall the DNN predicted calibration factors behave well with values close to the ATLAS standard calibration factors.

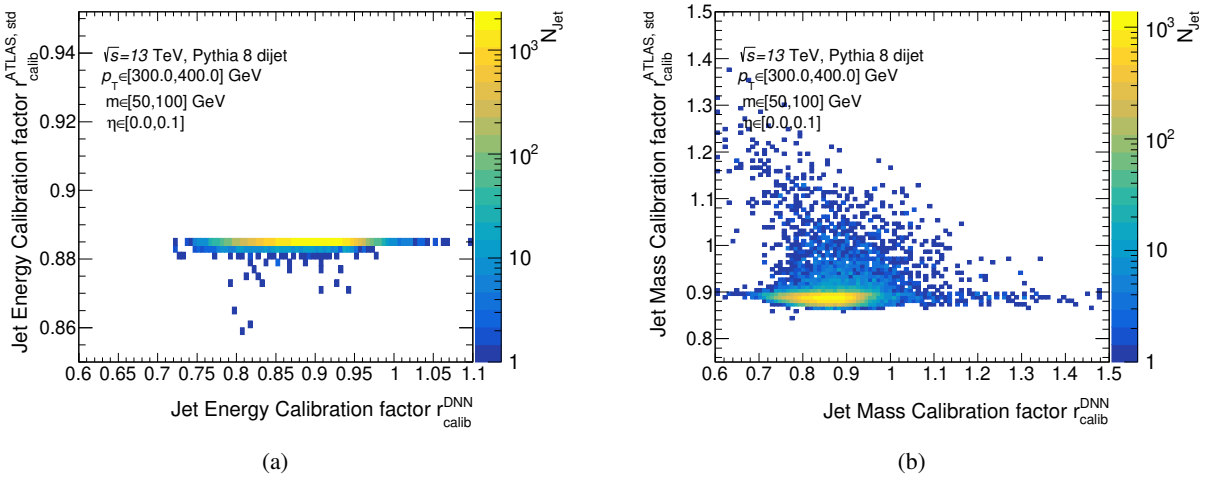


Figure 6.19: Jet energy (a) and mass (b) calibration factors predicted by the DNN ($r_{\text{calib}}^{\text{DNN}}$) versus provided by the standard calibration ($r_{\text{calib}}^{\text{ATLAS, std}}$) for one bin in $(p_T^{\text{true}}, m^{\text{true}}, \eta^{\text{true}})$.

¹² The ATLAS standard calibration is derived in bins of E^{true} and η^{true} , hence jets with similar energy and η will have similar calibration factors.

Dependency on the generator: The DNN calibration will be applied to all simulated samples used in ATLAS analyses that are produced with a variety of generators, as well as jets in real data events. Therefore, it is desirable to evaluate if the performance of the network is strongly affected by the modelling of jets within PYTHIA8. Generators have specific ways of evaluating the parton shower and hadronisation resulting in differences in the predicted particle flow and simulated detector showers. Since such differences can be representative of those between simulation and real data, it is important to verify they are not detrimental to the performance of the DNN calibration.

A data sample of SHERPA-generated dijets with the same data preparation as described in Section 6.2.2 is used to compute jet energy and mass responses with the standard calibration and the DNN calibration¹³. Both are then compared with the performance obtained with the PYTHIA dijet data sample. Figure 6.20 shows this comparison.

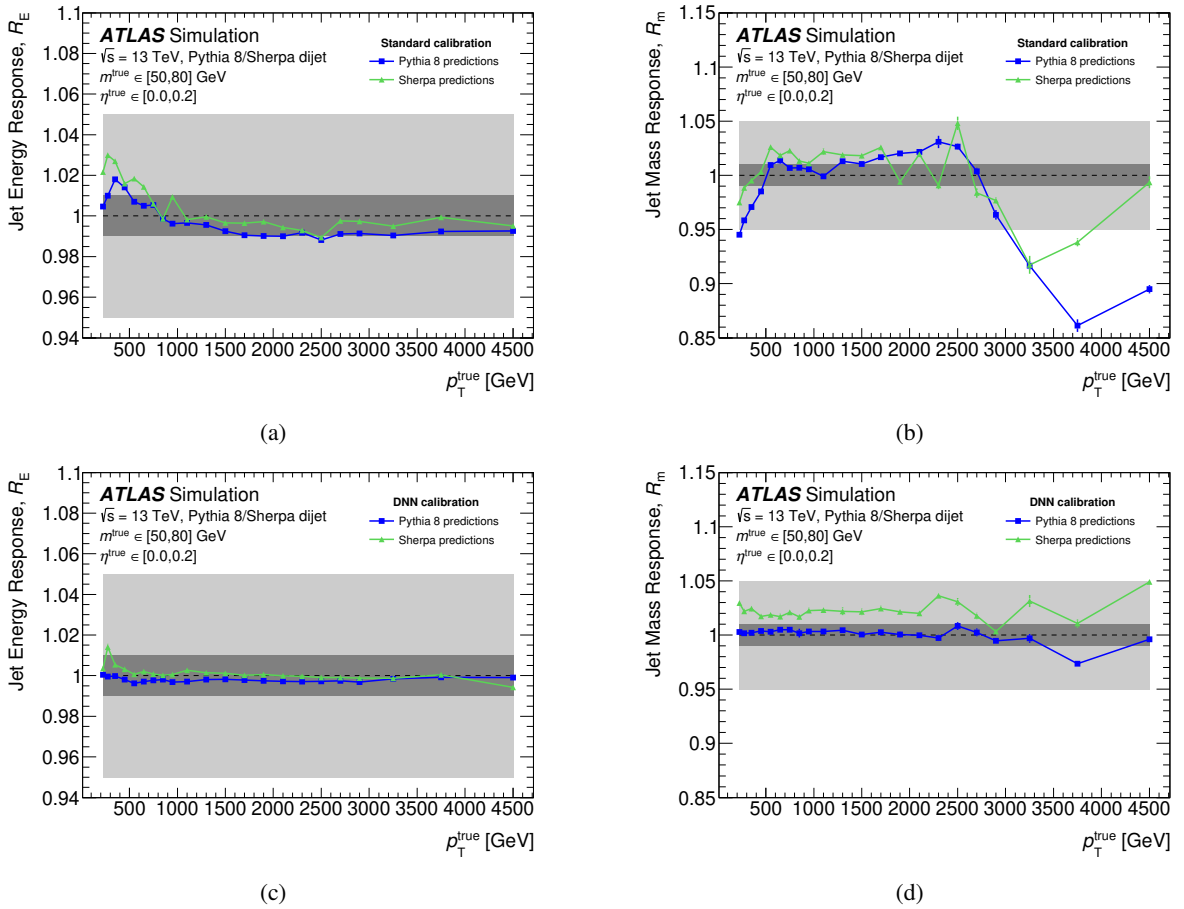


Figure 6.20: Jet energy and mass responses versus p_T^{true} calibrated with (a, b) the standard calibration and (c, d) with the DNN predictions for a dijet data sample generated with PYTHIA (squares) or SHERPA (triangles). The horizontal bands represent 1% (dark) and 5% (light) up and down divergences from closure.

The DNN energy calibration shows almost identical performance for the two generators, except at low p_T^{true} where a small under-performance is observed with the SHERPA data sample, although closure is however still obtained. Moreover, the DNN calibration shows a lesser dependency on the generator than the standard

¹³ The DNN is trained with the PYTHIA sample as discussed in the previous sections.

calibration for the energy. For the mass calibration with the SHERPA data sample, the DNN response closes within 5% but outside 1% in all the p_T^{true} range. This can be expected as the jet's mass is more dependent on its particle flow structure. This under-performance is observed to be reduced in higher mass bins. The standard mass calibration shows smaller differences between the two generators but overall, the DNN mass calibration is still better than the standard one for both the PYTHIA and SHERPA generators. Moreover, the residual non-closure that could appear for jets in real data will in any case be corrected for by the in-situ calibration step.

6.6 Limitations and possible solutions

Despite the satisfying results obtained with the calibration methodology presented above, some limitations and open questions were identified. They are discussed together with possible solutions in this section.

6.6.1 Training convergence and ending criterion

A weakness in this DNN training procedure is the lack of a clear criterion to mark the final convergence of the network and end the training. It is observed that early interruptions of the training (as soon as the losses start to fluctuate around a minimum) result in worse calibration performance in terms of JES, JMS or resolution in significant parts of the phase space. Instead, training for numerous additional epochs (such as described in Section 6.4.4) during which the training loss remains essentially constant is found to be necessary. A possible explanation is that the goodness of the performance is based on relatively strict expectations (e.g., closure within 1% for the energy response) to be obtained over all the (E, η, m) phase space while only a part of it is statistically dominant in the training sample. As a consequence, the global loss can be mostly determined by this dominant phase space region in which the optimum is quickly reached, while it requires more training to improve the sparsely populated regions. This behaviour also resembles the ‘grokking phenomenon’ observed in some DNN problems where sometimes performance suddenly increases after a very long training period [117].

Although the present study shows that training long enough allows coverage of the full phase space, a robust methodology would require a well-understood criterion to stop the training. Evaluating the calibration performance across all the phase space is not a practical option as it is too costly to perform automatically at each epoch. An interesting approach would be to evaluate the loss in predefined, well-chosen bins of the phase space (in terms of statistical contents and of physics relevance). Monitoring the losses in these bins with training and validation samples would possibly be more indicative than monitoring the virtually flat total loss.

6.6.2 Calibration of heavy boosted jets

During the validation of the DNN calibration performance for W/Z , Higgs boson and top-quark jets, non-closures of up to a few percent are observed for the energy (see Figure 6.13). While this under-performance is also visible for the standard calibration, the DNN calibration performance is slightly worse in these cases. Contrary to the DNN calibration presented here, the standard calibration is obtained using different types of jets. At high p_T (above ~ 500 GeV) QCD jets are used, while at low p_T the calibration is only derived from W/Z jets as this choice leads to better calibration performance for both W/Z jets and QCD jets.

A similar approach with mixed training samples is possible with the DNN. However, simply adding W/Z samples to the training does not improve – or even degrades – the performance for mass regions away from the W and Z masses. This issue could be mitigated with the availability of special pseudo- W/Z samples having a flat mass distribution. Alternatively, an ad hoc solution mimicking the standard calibration is possible by adding a subnetwork to the already QCD-trained DNN. This subnetwork would branch from the input layer with η annotation, produce output values used as multiplicative correction factors to the main network predictions and enforce these factors to be 1.0 for large p_T jets (typically above 500 GeV) to maintain the good performance of the DNN in this phase space. The subnetwork could then be trained on

W/Z samples, while keeping the main network frozen to improve the lower p_T regions. This is resumed in Figure 6.21.

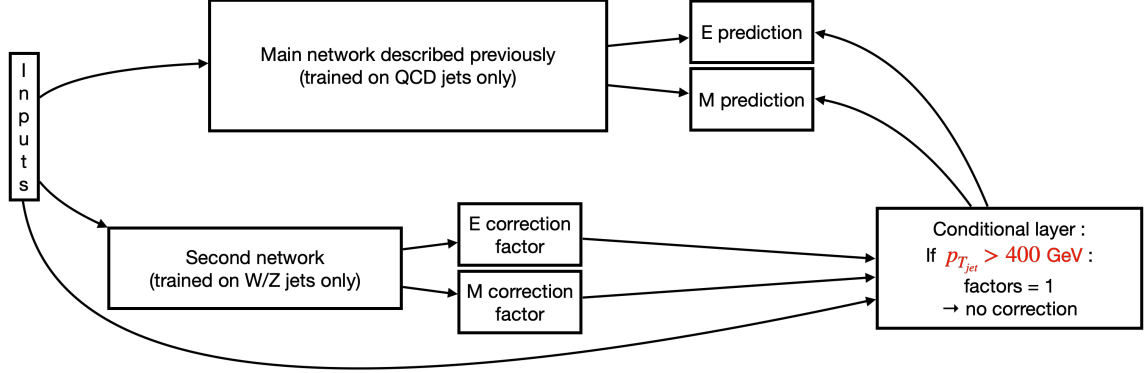


Figure 6.21: Diagram of the subnetwork principle.

This procedure allows to leave the main network untouched while applying a correction to the energy and mass predictions of low- p_T jets that is based on a W/Z -jet training. This is currently under investigation.

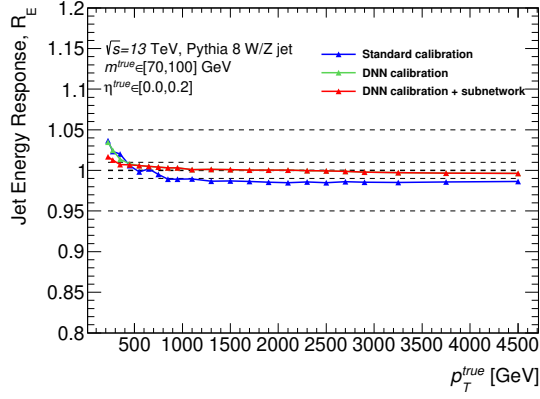
Table 6.4 gives a possible extended training procedure that was put in place to test this method. The two additional steps (12 and 13) are the subnetwork-only training on a W/Z sample, all the main-network weights being frozen.

Table 6.4: Extended training procedure.

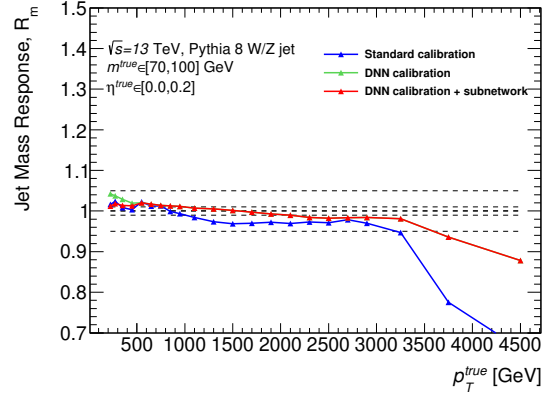
Steps	N°	Number of epochs	Batch size	Loss
...	...	...	...	...
Exclusive mass training	11	50	95000	MDN truncated (1.0σ)
Subnetwork training, W/Z samples	12	20	95000	MDNA truncated (E: 3.0σ , m: 2.0σ)
	13	20	95000	E: MDNA truncated (3.0σ), m: MDN truncated (1.0σ)

With this method, improvements in the DNN predicted calibration are observed for W/Z jets in the low- p_T region and for the W/Z mass bin (70 – 100 GeV), while small performance changes are observed in this specific mass bin for QCD jets. Figure 6.22 shows the W/Z -jet energy and mass responses comparing the original DNN to the one described here and Figure 6.23 shows the results on QCD -jet.

However, some problems arise: these improvements are only seen in this specific mass bin and decreased performances are observed in other mass bins for the calibration of W/Z -jets. The main reason is thought to be the use of the W/Z -jet sample to train the subnetwork. This sample is largely biased in its mass distribution which peaks at around 70 – 100 GeV. Thus, training with this sample creates a bias in the prediction of the correction factors: the high-statistics mass bin is well predicted while the other mass regions are not. In order to test this hypothesis and counter this effect, the use of a flat-mass W/Z -jet sample is planned for the training. This flat-mass sample could also potentially lead to improvements of the DNN calibration for top jets and Higgs jets.

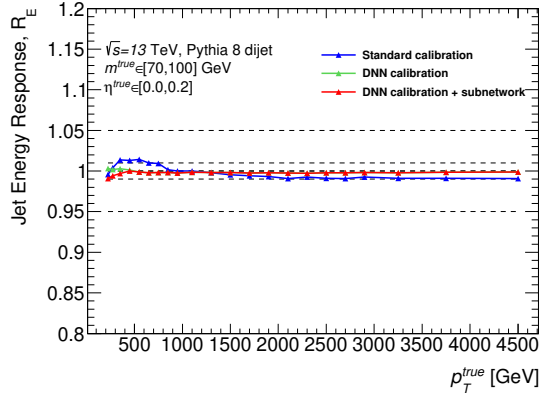


(a)

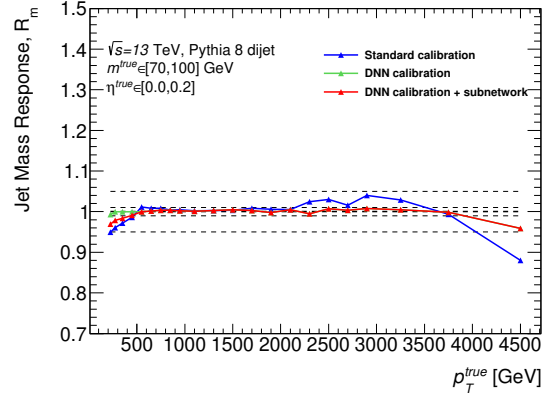


(b)

Figure 6.22: W/Z-jet energy (a) and mass (b) response for different calibrations (blue: ATLAS standard calibration, green: DNN predicted calibration, red: DNN predicted calibration with subnetwork) as a function of p_T^{true} . The dotted lines represent 1% and 5% up and down divergences from closure.



(a)



(b)

Figure 6.23: QCD-jet energy (a) and mass (b) response for different calibrations (blue: ATLAS standard calibration, green: DNN predicted calibration, red: DNN predicted calibration with subnetwork) as a function of p_T^{true} . The dotted lines represent 1% and 5% up and down divergences from closure.

6.6.3 Importance of the features

Insights into which input features are the most influential on the DNN predictions are important for two reasons. Firstly, they generally help to understand the jet energy and mass responses. Secondly, they can indicate if some features could be removed from the inputs, potentially leading to a simpler DNN or even to better performance if the removed variables are confusing the network. A simple way to quantify the importance of a feature is to measure the average of the relative variation in the prediction when this feature is systematically varied for each input jet. A ranking of the different features obtained by following this method is shown in Figure 6.24 where each feature is varied by $\pm 1\%$ of the standard deviation of its overall distribution. Both the energy and mass prediction variations are displayed for each feature.

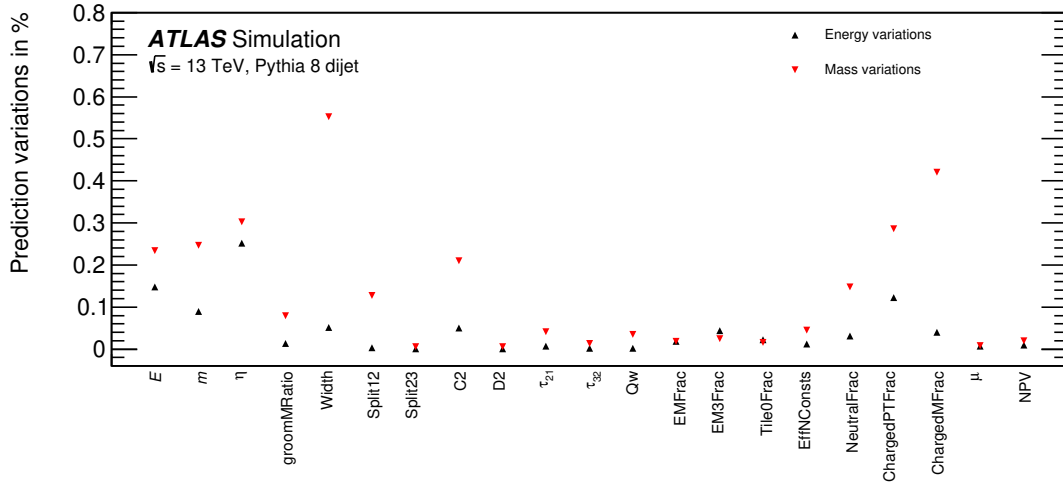


Figure 6.24: The variation of the energy (upward triangles) and mass (downward triangles) predictions when input features are varied.

As expected, the variables E , η , and m have a large influence on the predictions according to this measure. Some substructure features¹⁴ have a stronger impact on the mass prediction than on the energy prediction. This is also expected since the mass depends on the structure of the particle flow objects in the jet. While by this metric μ and NPV do not seem so important, it was demonstrated in Section 6.5.3 that these features are important as they allow the DNN calibration to remove the response dependence on these variables. Further investigations are therefore necessary to understand more precisely the importance of the features. Those could consist of considering the magnitude of the DNN gradient with respect to its inputs in bins of p_T or mass, or to apply the SHAP methodology.¹⁵

¹⁴ The Width, groomMRatio, ChargedPTFrac, ChargedMFrac, Split12, and C2.

¹⁵ SHapley Additive exPlanations [118].

6.7 Conclusion

A successful full implementation of a deep neural network based simultaneous calibration of the energy and mass of large- R jets with the ATLAS detector was presented in this Chapter. This implementation solved problems that are specific to the calibration task. Firstly, rather than segmenting the calibration in angular slices, a single network covers multiple detector regions and a dedicated encoding of the jet angular direction was included to help it to adapt to the corresponding strongly varying response. Secondly, the network architecture was designed to efficiently accommodate both the energy and the mass predictions. Finally, loss functions were carefully chosen so that the DNN prediction converges as desired toward the most probable value of the energy or mass response distribution; this also implies non-trivial training procedures involving variants of the loss functions.

The resulting calibration was compared to the standard calibration on QCD jets and on other jet topologies not seen during the DNN training. Over all the phase space and for all processes of interest, the DNN was found to achieve better energy and mass scale closure and similar or better resolution, with a typical 30% improvement in the energy resolution for $p_T > 500$ GeV. The DNN calibration was also shown to be robust against various tests such as pileup dependency or changes to the sample distribution. However, some limitations were acknowledge and possible improvements were discussed. Overall, as this method can directly replace the current simulation-based steps of the ATLAS large- R jet calibration chain, no adaptation of the data-based methodology is required and this DNN procedure can be used for the calibration of large-radius jets used in future ATLAS analyses. The DNN calibration for large- R UFO jets will be used for recent Run 2 analyses and for all the Run 3. For each data taking period, a dedicated DNN need to be trained on MC simulations emulating the corresponding ATLAS configuration, and LHC's energy and pileup profile.

I was the main contributor to this project under the supervision of Pierre-Antoine Delsart. My work spanned from designing the DNN architecture and training to the validation of the calibration performance. It was followed by the writing of a paper [90] for which I am the main author with Pierre-Antoine Delsart. This project was initially part of my qualification task to become an author of the ATLAS Collaboration but was ultimately a continued effort along my thesis in order to provide calibration recommendation to the Collaboration. Since the end of this qualification I provided new trained DNNs to be used for the latest reprocessed Run 2 data and early Run 3 data (2022+2023).

CHAPTER 7

SEARCH FOR EMERGING JETS WITHIN DARK QCD MODELS WITH THE ATLAS DETECTOR DURING RUN-3

Outline of This Chapter

7.1	Data and Monte Carlo simulations	113
7.2	Event reconstruction, objects selections and efficiencies	118
7.3	Trigger studies	126
7.4	Analysis strategy	131
7.5	Discriminating observables and data-MC agreement	135
7.6	Event selections: ABCD plane and signal region definitions	150
7.7	Background estimation, validation of the ABCD method	155
7.8	Systematic uncertainties	167
7.9	Statistical analysis and results	174
7.10	Conclusion	190

The second main project of my thesis was the development of an ATLAS Run 3 data analysis on the search for emerging jets (EJs) within dark QCD models in collaboration with a team of international analysers.

This analysis purpose is to search for BSM physics with the EJs phenomenology which was described in Chapter 3. As discussed, the dark QCD and EJs phenomenologies are smoking guns for BSM physics thanks to the highly exotic signature it produces in particle detectors and particularly, in the context of this thesis, in the ATLAS detector. The studied model is the one described in Section 3.3 with the production of two EJs through the decay of a new vector boson Z' . Several sets of model parameters are considered, varying the mass of the Z' and the dark pion mass and lifetime. This analysis is based on the data recorded by ATLAS during the two first years of the LHC's Run 3, 2022 and 2023, corresponding to an integrated luminosity $\mathcal{L}_{\text{int}}$ of 51.8 fb^{-1} .

Most of the analysis developments were made on MC simulations of both the EJ signal and the main source of background, which is QCD dijet events. The MC simulations allow to find highly discriminating observables between signal and background. These discriminating observables are then used to define a signal-enriched region in which the background is as limited as possible. The analysis strategy is to estimate the expected background contribution with a data-driven method in this signal region and compare

the estimation to data to state if there is an excess of events that could originate from BSM physics. If no excess is observed, exclusion limits on the EJ production cross-section would be set to exclude the tested model parameters. The objects of interest for this analysis are the jets and the associated tracks and vertices linked to the decays of the dark hadrons into visible decay products. A specific jet collection was used to accommodate for the specific analysis needs.

Within the analysis team I am one of the main contributors for a broad range of tasks. I heavily contributed to the event reconstruction step and object definitions (in particular for the jet collection), implementing these into an existing DAOD format (derivation data processing step). I am also one of the main maintainers of the analysis pipeline framework used to select and calibrate the objects of interest and compute the observables necessary for the analysis. Finally, I solely developed one of the analysis strategies from the trigger and signal-background discrimination studies to the signal region definition, background estimation method and statistical analysis.

This Chapter will present in details this analysis strategy and the obtained results. First, Section 7.1 describes the data and MC simulations exploited. Section 7.2 presents the physics objects of interest, how they are reconstructed and selected. The triggers considered in the analysis are described in Section 7.3. An overview of the analysis strategy is given in Section 7.4 followed by the presentation of: the discrimination studies (7.5), the event selections and signal region definition (7.6) and the background estimation (7.7). The systematic uncertainties impacting the analysis are introduced in Section 7.8. Finally, the statistical analysis and partial results are presented in Section 7.9. As this analysis is still under approval within the ATLAS Collaboration at the moment of writing this thesis, only data-blinded results will be discussed, along with expected sensitivities computed from MC simulations.

7.1 Data and Monte Carlo simulations

7.1.1 Data

This analysis uses data events of pp collisions at $\sqrt{s} = 13.6$ TeV collected by ATLAS from 2022 and 2023 during the beginning of the LHC's Run 3 for a total integrated luminosity $\mathcal{L}_{\text{int}} = 51.8\text{fb}^{-1}$. The per-year integrated luminosities are: $\mathcal{L}_{\text{int}}^{2022} = 26.1\text{fb}^{-1}$, $\mathcal{L}_{\text{int}}^{2023} = 25.7\text{fb}^{-1}$. Data events undergo some quality filtering to remove events recorded when the LHC or ATLAS detector were not running optimally (e.g. a non-working sub-detector) [59].

7.1.2 Monte Carlo simulations: signal and background

As firstly discussed in Section 4.4, MC simulations of the theoretical EJ signal and of background events are used in this analysis to study the characteristics of the signal and in particular their specificities regarding the background events. This allows to design the selection criteria to apply in order to discriminate signal from background and to define a region of phase space where the signal events are dominant. Moreover, the MC samples allow to design and test the analysis strategy and prepare the statistical interpretation. The MC simulations used in this analysis are part of the MC23 simulation campaign of ATLAS and are representative of the Run 3 LHC's energy and pileup conditions and Run 3 ATLAS detector's configuration.

7.1.2.1 Background samples

In this analysis, the EJ signal signature is composed of two dark jets in the final state, thus the dominant background is by far composed of dijet events which are events containing two SM QCD jets than can mimic the signal in particular cases. As in the previous Chapter, the dijet MC sample is generated using PYTHIA 8.230 [60] with the NNPDF2.3LO [93] parton distribution function (PDF) set and the A14 set of tuned parameters [94]. This sample is composed of jets originating predominantly from light quarks and gluons. Jets originating from bottom or top quarks can be generated but are essentially drowned in the light quark- and gluon-jet production. The dijet events are generated in slices of p_T , to sufficiently populate the kinematic region of interest (from around 200 GeV up to 6000 GeV).

In addition, to study the efficiencies of the jet triggers used in the analysis and the agreement between MC simulations and data for jet observables, a control region is defined based on a single-muon trigger (this choice of control region will be motivated in Section 7.3). Such a selection, and objective, motivates the need for additional samples containing both jets and muons. The three main processes that can produce such a combination are:

- $W \rightarrow \mu\nu_\mu + \text{jets}$,
- $W \rightarrow \tau\nu_\tau + \text{jets}$, when the τ decays to $\nu_\tau\mu\nu_\mu$,
- $t\bar{t}$, when one of the top decays hadronically ($t \rightarrow Wb \rightarrow q\bar{q}b$) and the other leptonically ($t \rightarrow Wb \rightarrow l\nu_l b$ where l is a muon).

The W +jets samples are generated using SHERPA 2.2.1 [61] and next-to-leading-order (NLO) matrix elements (ME) for up to two jets, and leading-order (LO) matrix elements for up to four jets calculated with the Comix [119] and OPENLOOPS [120–122] libraries. They were matched with the SHERPA parton shower [96] using the MEPS@NLO prescription [123–126] using the set of tuned parameters developed by the SHERPA authors. The NNPDF3.0_{NNLO} set of PDFs [127] was used and the samples were normalised to a next-to-next-to-leading-order (NNLO) prediction [128].

The $t\bar{t}$ sample is generated in three different sub-samples corresponding to the different decays of the tops: two hadronic decays (referred as *all hadronic*), one leptonic decay + one hadronic decay (referred as *single leptonic*) and two leptonic decays (referred as *di-leptonic*). The production is modelled using the POWHEG BOX v2 [129–132] generator at NLO with the NNPDF3.0_{NLO} [127] PDF set. The events are interfaced to PYTHIA 8.230 to model the parton shower, hadronisation, and underlying event, with parameters set according to the A14 tune and using the NNPDF2.3_{LO} set of PDFs. The decays of bottom and charm hadrons are performed by EVTGEN 1.6.0 [133]. The $t\bar{t}$ samples are also considered as a sub-dominant source of background for the analysis.

Table 7.1 summarizes the different MC background samples.

Table 7.1: MC simulated background samples

Background	Use	Generator
QCD dijet	Dominant background	PYTHIA 8.230
W +jets	Muon control region	SHERPA 2.2.1
$t\bar{t}$	Muon control region + sub-dominant background	POWHEG BOX v2 +PYTHIA 8.230

7.1.2.2 Signal samples

The signal samples are generated with PYTHIA 8.230 through the Hidden Valley module [134]. This module is an extension of PYTHIA which adds a simplified dark QCD sector and the processes to access it. It adds a Z' vector mediator as described in Section 3.3 which is produced through a quark-antiquark annihilation and decays to a pair of SM quarks or dark quarks with configurable branching ratios. The additional dark QCD sector presents a simplified phenomenology with free configurable parameters. It contains N_f flavours of dark quarks, that are, however, degenerate in mass and differ only by a dark sector flavour, and each are charged under the dark QCD gauge group with N_{CDM} colours. Only the two lightest dark mesons of spin-0 (dark pion π_D) and spin-1 (dark rho ρ_D) are simulated with the supposition that the heavier dark mesons decays promptly to the lightest ones. If kinematically allowed the dark rho decays promptly to a pair of dark pions and the dark pion decays to a pair of SM quarks with a lifetime $c\tau_{\pi_D}$. Dark baryons are not simulated as they are not yet implemented in the Hidden Valley module, it is not seen as a downside as dark meson production should dominate the dark baryon production for $N_{\text{CDM}} = 3$ as in the SM. Moreover, it is supposed that the lightest dark baryon is stable and will escape the detector without interacting, thus not impacting the detector measurements other than in the form of missing transverse momentum (MET) aligned with the pair of EJs. Finally, the dark QCD confinement scale Λ_D defines the overall energy scale of the dark sector.

The free and configurable model parameters in the Hidden Valley module are different, but equivalent, from the parameters presented in Section 3.3:

- $m_{Z'}$,

- $BR_{Z' \rightarrow q\bar{q}}$ and $BR_{Z' \rightarrow q_D \bar{q}_D}$, the branching ratios of the Z' to a pair of SM quarks and a pair of dark quarks (the branching ratios are directly related to the coupling constants of the Z'),
- N_f, N_{CDM} ,
- Λ_D ,
- (x, y, z) , the factors fixing the masses of the dark particles, $m_{q_D} = x\Lambda_D$, $m_{\pi_D} = y\Lambda_D$, $m_{\rho_D} = z\Lambda_D$,
- $c\tau_{\pi_D}$.

For all generated signal samples the number of flavours N_f is set to 7 and the number of colours N_{CDM} to 3, no variations of these two parameters are considered as it is supposed that the dark QCD sector dynamics does not impact much the fundamental EJs signature. Moreover, variations of N_f and N_{CDM} would produce fluctuations in the dynamic of the dark sector's parton shower and hadronisation mainly resulting in variations in the number of dark radiations and average amount of produced dark hadrons. These kinds of fluctuations are already considered through the variations of the dark particles and mediator masses ($m_{q_D}, m_{\pi_D}, m_{\rho_D}, m_{Z'}$). The dark particle masses' relations to Λ_D are taken as such: $m_{q_D} = \Lambda_D$, $m_{\pi_D} = 0.5\Lambda_D$ and $m_{\rho_D} = 2\Lambda_D$. The Z' branching ratios are chosen to have the following values, $BR_{Z' \rightarrow q\bar{q}} = 0.001$ and $BR_{Z' \rightarrow q_D \bar{q}_D} = 1 - BR_{Z' \rightarrow q\bar{q}} = 0.999$ such that most of the Z' decay to dark quarks and EJs but also to limit the cross-section of $Z' \rightarrow q\bar{q}$ which is strongly constrained by searches for dijet resonances [135–137]. Three signal models, A, C and D, are considered with different Λ_D values and, as such, different dark pion and dark rho masses. For each signal model two dark pion lifetimes are considered (5 and 50 mm) and three Z' masses (600, 1500 and 3000 GeV) for a total of 18 signal samples. This is resumed in Table 7.2.

Table 7.2: MC simulated signal samples

	Model A	Model C	Model D
Λ_D (GeV)	10	20	40
m_{π_D} (GeV)	5	10	20
m_{ρ_D} (GeV)	20	40	80
$c\tau_{\pi_D}$ (mm)	5-50		
$m_{Z'}$ (GeV)	600-1500-3000		
Decay to SM	$\pi_D \rightarrow d\bar{d}$		
Decay in dark sector	$\rho_D \rightarrow \pi_D \pi_D$		
Number of generated events	200,000		

In the rest of the analysis the different signal samples will be denoted according to the following structure:

$$\text{Ld}\{\Lambda_D\}_{\text{rho}\{m_{\rho_D}\}_{\text{pi}\{m_{\pi_D}\}_{\text{Zp}\{m_{Z'}\}_{\text{l}\{c\tau_{\pi_D}\}}$$

For example, for Model A with $c\tau_{\pi_D} = 5$ mm and $m_{Z'} = 1500$ GeV the sample name is:

$$\text{Ld10_rho20_pi5_Zp1500_15}$$

7.1.2.3 Pileup overlay

In addition, for all MC samples, the effect of pileup was modelled by overlaying the simulated hard-scattering events with inelastic pp events generated with PYTHIA 8.186 [99] using the NNPDF2.3LO set of PDFs [93] and the A3 set of tuned parameters [100]. The number of overlaid events is representative of the amount of pileup present in the 2022–2023 ATLAS pp data samples (two first years of Run 3). However, as the pileup conditions have quite differed between these two years, as shown in Fig. 7.1, each background and signal sample is generated twice: one sample matches the 2022 pileup conditions (referred as MC23a) and the other matches the 2023 pileup conditions (referred as MC23d). Both samples are then combined to match the overall pileup conditions of both years.

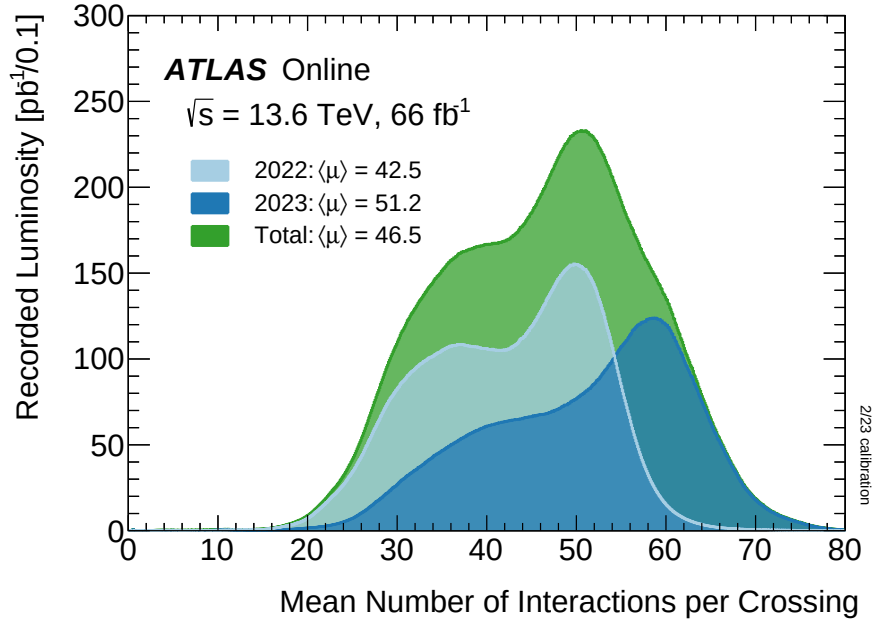


Figure 7.1: Recorded integrated luminosity as a function of the mean number of interactions per bunch crossing with ATLAS during the two first years of Run 3: 2022 (light blue), 2023 (dark blue) and combination (green) [58].

7.1.2.4 Monte Carlo re-weighting

In order for a specific MC sample to be representative of the real data in terms of number of events and observable distributions (e.g. jet p_T , muon η , ...), each simulated event is associated to a normalisation weight w_{tot} which is used when constructing histograms. This weight depends on the MC simulation properties, the cross-section of the simulated process, and the data parameters it should represent (e.g. $\mathcal{L}_{\text{int}}$). The definition of w_{tot} is given in Eq. 7.1:

$$w_{\text{tot}} = \frac{w_{\text{MC}}}{\sum w_{\text{MC}}} \times f_{\text{eff}} k \sigma \times \mathcal{L}_{\text{int}} \times w_{\text{pileup}} \quad (7.1)$$

where:

- w_{MC} is the MC event weight defined by the generator,

- $\sum w_{\text{MC}}$ is the total sum of the individual event weights of the MC sample, it is used to normalise the number of events such that the weighted number of events ($\sum w_{\text{tot}}$) in the samples does not depend on the generated MC statistics but only on the cross-section and luminosity,
- f_{eff} is the event filter efficiency of the generator, it corresponds to the fraction of generated events included in the MC sample,
- k , referred as the k factor, is used to take into account effect from higher order of perturbative theory in the generator cross-section computation,
- σ , the total cross-section of the simulated process, computed by the generator, this number is corrected by the filter efficiency f_{eff} and k factor in order to be coherent with the MC actual composition,
- $\mathcal{L}_{\text{int}}$, the integrated luminosity of the data that the MC simulation should represent,
- w_{pileup} is a weight related to the pileup re-weighting, it is used to match the pileup profile in terms of actual μ distribution between the MC sample and data.

For a particular process ($pp \rightarrow \text{final state}$), the following equality should be approximately true:

$$N_{\text{weighed event}}^{\text{MC}} \equiv \sum w_{\text{tot}} = N_{\text{event}}^{\text{data}}$$

For re-weighting and combining the two MC sub-campaigns (MC23a and MC23d) the integrated luminosity of 2022 and 2023 data-taking years are used,

$$\mathcal{L}_{\text{int}}^{2022} = 26.1\text{fb}^{-1}; \mathcal{L}_{\text{int}}^{2023} = 25.7\text{fb}^{-1},$$

along specific pileup weights corresponding to the actual μ distributions of both years.

7.2 Event reconstruction, objects selections and efficiencies

All data and MC events are reconstructed following the procedure discussed in Section 5.1 and are required to be selected by at least one of the analysis triggers (muon or jet triggers) as described in the next section. The specificities of the analysis in terms of object reconstructions and selections are discussed in this Section and are taken into account with a special derivation called LLP1 specifically developed towards long-lived particles (LLP) searches.

In this analysis the objects of interests are the jets, tracks and secondary vertices. Additionally, standard reconstructed isolated photons are used in an overlap removal (OR) procedure to veto out jets which contain a tightly isolated photon. This OR procedure allows to avoid events in which a photon has been falsely reconstructed as a jet. Moreover, reconstructed tightly isolated muons are used to define a muon control region in order to study the jet triggers efficiencies and the agreement between data and MCs. Finally, the hard scatter (HS) vertex is selected as the reconstructed primary vertex with the largest squared sum of the p_T of the associated tracks.

7.2.1 Jets

Due to the exotic signature of the EJs, the choices of the angular opening parameter R and of the constituents used to reconstruct jets in this analysis need to be carefully made in order to capture as well as possible the EJ specificities. These choices are however partly constrained by technical aspects imposed by the ATLAS Collaboration.

First on the angular opening, the process of production of the EJ where dark pions decay to SM quarks leads to visible jets in which two hadronisation occurred (first in the dark sector and then in the SM sector) making EJ significantly broader than SM jets. The actual angular opening of the EJ is dependent on the mass of the Z' and the mass of the π_D . These two masses interplay in the boost given by the Z' to the π_D and then to the SM hadrons, thus determining the final collimation of the jet constituents. The lighter the masses of the Z' and π_D the broader the EJ are but the exact dependence is more complex to determine. In the ATLAS experiment, the standard jet angular opening for broad jet is the one of the large- R jet with $R = 1.0$. The use of larger jet can lead to difficulties, especially in terms of pileup contamination. Hence, for this analysis it was decided to use the standard large- R jet reconstruction with $R = 1.0$. However, a particular choice of constituents was made to partially take into account scenarios where the EJs would not be completely contained in a $R = 1.0$ jet.

Second on the jet constituents, at the beginning of this analysis there were two standard constituents for large- R jets in ATLAS:

- LCTopo with the Trimming procedure applied (see Sec. 5.4.4),
- UFO with the CHS, CSSK and SoftDrop procedures applied (see Sec. 5.4.4).

Both were problematic for this analysis, LCTopo for a technical reason and UFO for a physical reason.

The LCTopo large- R jets are on the paper completely adequate for this analysis as LCTopo constituents are performant to reconstruct hadronic jets thanks to the LCW cluster calibration. However, since the beginning of Run 3, LCTopo jets are deprecated for physics analyses in favour of the more recent and more performant UFO constituents (as discussed in Section 5.4.3). In the ATLAS Collaboration this means that this collection is not supported by the performance group which provides the calibrations and

uncertainties for jets. To use this collection would thus require to develop internally to the analysis the necessary calibration and uncertainties. As this is a heavy task which would only be beneficial for this analysis, it was decided to not use this collection, especially as another satisfactory solution was found, as described below.

On the other hand, UFO large- R jet are supported by the ATLAS Collaboration and calibration and uncertainties are thus provided for all physics analyses. However, contrary to the LCTopo constituents the UFO constituents are not optimal for reconstructing EJs. UFO constituents use a combination of clusters and tracks and the CHS procedure discard any charged constituents that is not originating from the HS vertex. In the case of LLPs, like for the EJs, this is very problematic as a large portion of the tracks in the jets will originate from a LLP decay and thus not from the HS. This leads to losses in the energy reconstruction of the jets using UFO constituents. For this reason the UFO large- R jets were also discarded.

As many LLPs searches aiming at using Run 3 data also had this problem with UFO constituents and more generally with the track-based constituents (PFlow, TCC), a choice of collection of jets that is optimal to all the LLP analyses was made and this collection is now supported by the LLP community and jet performance group in ATLAS. This collection is the EMTopo small- R jets (as many analyses needs small- R jets instead of large- R jets). For this analysis, the third possible choice was then to create a new collection of large- R jets by reclustering the EMTopo small- R jets called *RCEMTopo*. The principle of reclustering is to use the small- R jets themselves as the constituents to be clustered by the jet algorithm (here with $R = 1.0$). The 4-momentum of the RCEMTopo is thus the sum of its small- R jet constituents:

$$p_4^{\text{RCEMTopo}} = \sum p_4^{\text{EMTopo}}$$

This has two main advantages:

- as the RCEMTopo are constructed from the small- R jets, the calibration and uncertainties of the small- R jets can directly be propagated to the RCEMTopo jets, no additional calibration or uncertainties are needed,
- it can allow to capture better the EJ particles as the clustered particles can now extend to higher radius than 1.0 if one of the constituent small- R jet is positioned far from the centre of the RCEMTopo jet.

The collection of RCEMTopo large- R jets thus seemed the more optimal for this analysis.

To confirm that the collection of RCEMTopo jets can be used to optimally reconstruct EJs, a study was performed on the signal samples comparing the three mentioned jet collections. The distributions of the 2-jet invariant mass m_{jj} ¹ and of the leading- p_T jet energy response are compared for different signal samples. The three collections of jets are calibrated using Run 2 based calibrations as they were the only calibrations available at the beginning of Run 3. The results obtained for the samples Ld20_rho40_pi10_Zp1500_15 and Ld20_rho40_pi10_Zp1500_150 are shown in Fig. 7.2. For all signal samples, the RCEMTopo jet collection shows the best performance in reconstructing EJs both in terms of invariant mass (with a distribution peak closer to $m_{Z'}$) and energy response (with a distribution better centred on one). These results shown in Fig. 7.2 also presents clearly the effect of CHS discussed earlier for the UFO constituents, a significant decline in performance is observed between the sample with $c\tau_{\pi_D}$ of 5 and 50 mm, as more charged constituents are removed in the latter. From these results, it was decided to use RCEMTopo large- R jets for this analysis.

¹ The mass of the 4-momentum sum of the two leading- p_T jets. By momentum conservation, m_{jj} should be equal to $m_{Z'}$ with perfect reconstruction of the jets.

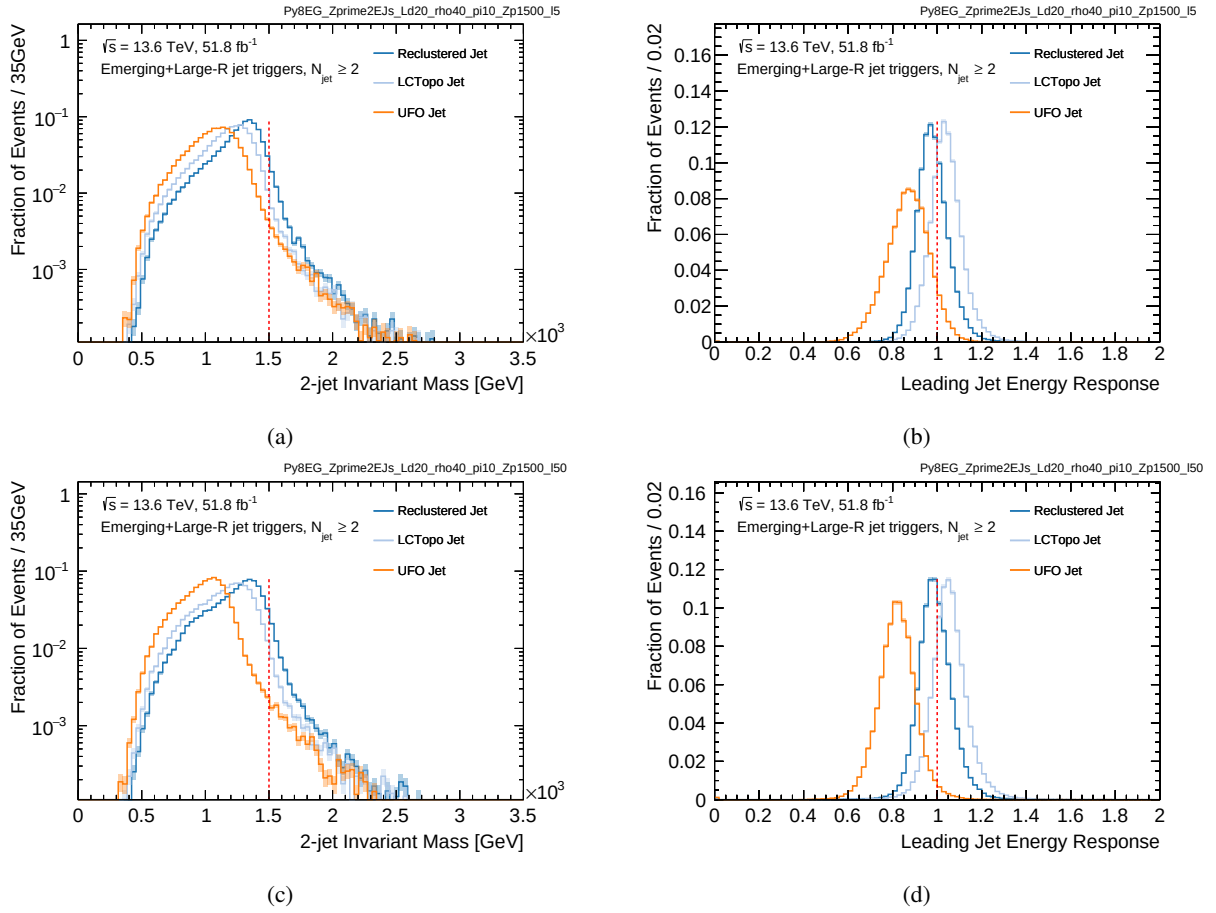


Figure 7.2: Comparison of the 2-jet invariant mass m_{jj} (a,c) and leading- p_T jet energy response (b,d) distributions reconstructed with three jet collections (RCEMTopo, LCTopo and UFO) for two signal samples, Ld20_rho40_pi10_Zp1500_15 (a,b) and Ld20_rho40_pi10_Zp1500_150 (c,d).

The reconstruction of this specific jet collection is added in the derivation LLP1 where a first pre-selection (standard for large- R jet) is applied:

$$p_T > 200 \text{ GeV} ; |\eta| < 2.5$$

At the derivation level, the RCEMTopo jets are constructed based on the small- R jet calibrated collection, this means that the RCEMTopo are calibrated by default. However, for flexibility it was necessary to be able to apply the jet calibration based on the DAOD files, within the analysis framework, and with the latest small- R jet calibration recommendation. At the DAOD level there are two possible procedures to calibrate reclustered jets:

- the exact procedure, which consists in applying the wanted small- R jet calibration to obtain the corresponding calibrated collection and re-apply the clustering based on this new calibrated collection in order to construct the RCEMTopo calibrated collection,
- the approximated procedure, which considers that the small- R jet calibration does not change the clustering and it is thus sufficient to recalculate the RCEMTopo jets 4-momenta using the one-to-one

corresponding calibrated small- R jet constituents without reclustering them:

$$\begin{aligned} \text{original 4-momentum: } p_4^{\text{RCEMTopo}} &= \sum_i p_4^{\text{EMtopo},i} \\ \text{apply small-}R \text{ jet calibration: } p_4^{\text{EMtopo},i} &\rightarrow p_4^{\text{EMtopo,cal},i} \\ \text{recalculate 4-momentum: } p_4^{\text{RCEMTopo,cal}} &= \sum_i p_4^{\text{EMtopo,cal},i} \end{aligned}$$

The exact procedure is, as the name suggests, an exact calibration per definition. But to re-apply the clustering is a CPU-intensive task and it is not expected that the calibration of the small- R jets has a significant impact on the clustering. To avoid useless calculations, the approximated calibration was thus favoured for its simplicity. To ensure that this approximation is correct a study was performed on several signal samples, a subset of dijet background (~ 1 million events) and a subset of data from 2022 ($\sim 50,000$ events) comparing for the leading- p_T jet the obtained kinematics differences between both applied calibration procedures. The results of this study are presented in Fig. 7.3 with the relative differences obtained for the p_T , mass, η and ϕ .

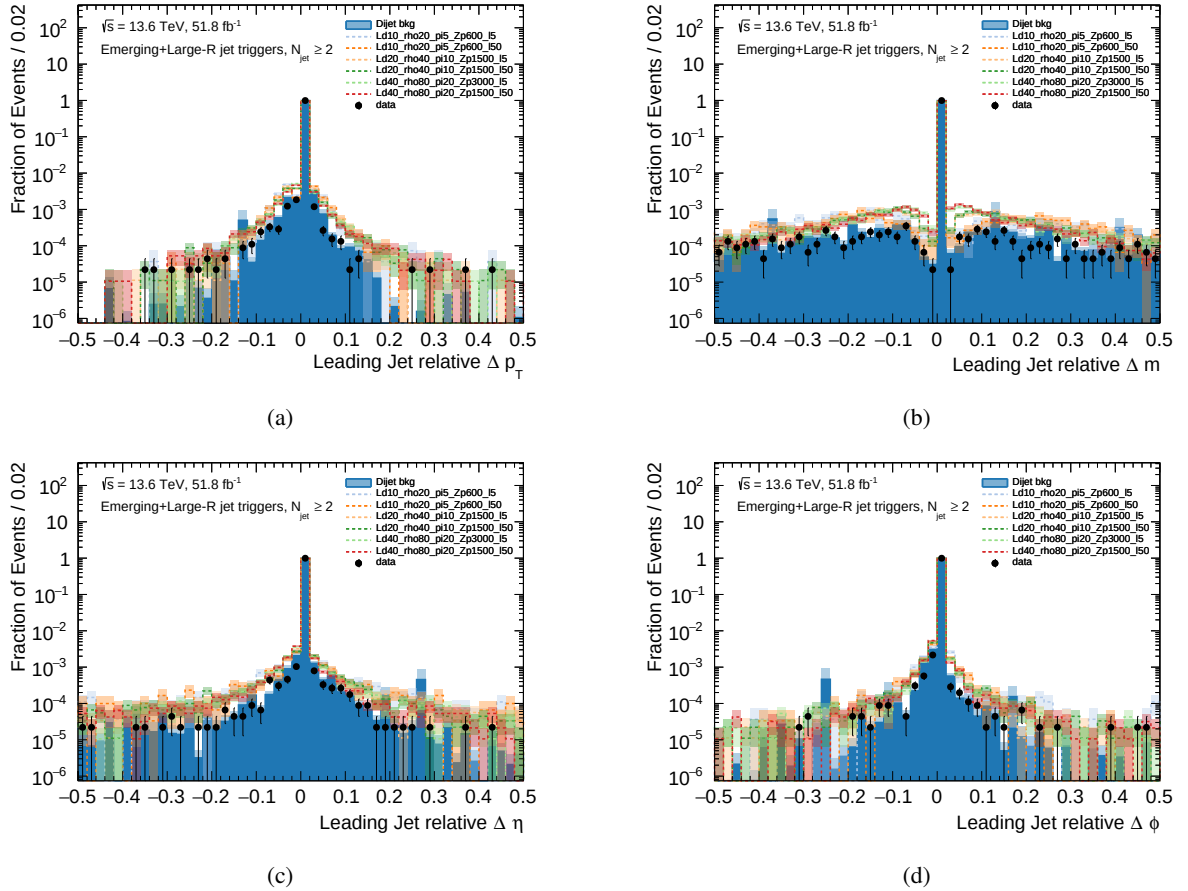


Figure 7.3: Relative differences for kinematics observables between the exact and approximated calibration procedures for RCEMTopo jets obtained on the leading- p_T jet for MC signal, MC dijet background and data.

It shows very limited differences between both calibrations with most of the distributions' statistics concentrated at zero and the rest of the distribution between three or four orders of magnitude lower. The approximated calibration was thus chosen for the analysis.

After the calibration, all the jets in an event are ordered in descending p_T and only the two most energetic jets, referred as leading and sub-leading jets, are considered in the analysis.

7.2.2 Tracks and secondary vertices

Tracks, both prompt and LRT, and secondary vertices are reconstructed following the procedure discussed in Sections 5.2 and 5.3. Only tracks and vertices that are within a RCEMTopo jets (i.e. $\Delta R < 1.0$ between the tracks/vertices and the jet) are considered in this analysis. All considered tracks are selected with $p_T > 0.5$ GeV (1.0 GeV for LRT tracks) and within the tracker acceptance $|\eta| < 2.5$. The secondary vertices are selected with an invariant mass $m > 0.6$ GeV and number of associated tracks $N_{\text{trk}} > 2$. These selections were defined based on the study of the secondary vertices reconstructed in MC signal and dijet background events in order to remove SM background vertices. Fig. 7.4 shows the distributions of the mass and N_{trk} of secondary vertices in MC. For the mass, a dominant peak in the dijet sample and less dominant in signals around 0.5 GeV is observed, this peak corresponds to the decay of the SM Kaon mesons which have a mass of 0.494 GeV for the charged Kaons and of 0.498 GeV for the neutral Kaon. This large source of background motivates the selection on the vertex mass at 0.6 GeV. For the number of tracks, about 70% of the dijet event vertices are composed of only two tracks, these kinds of vertices can for example be produced during photon conversions creating an electron-positron pair. To mitigate the number of vertices in the dijet events, the selection $N_{\text{trk}} > 2$ is added.

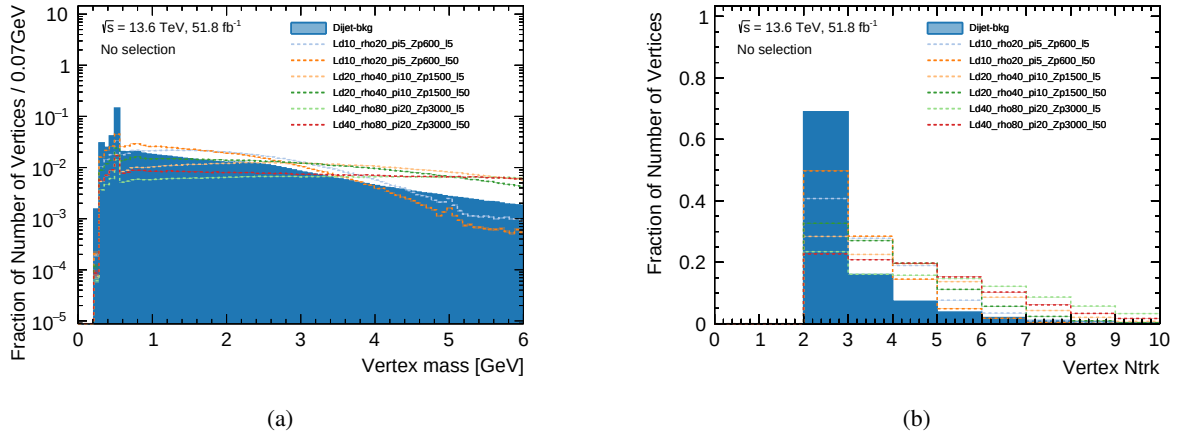


Figure 7.4: Distributions of the secondary vertices mass and number of associated tracks in MC signal and dijet background events.

In addition, background vertices can be produced during interactions of SM particles with the inner detector material. These interactions can produce reconstructed vertices located inside detector material. This phenomenon is clearly visible when constructing a plot of the x and y coordinates of vertices with $N_{\text{trk}} > 2$. This is presented in Fig. 7.5 with data events where the structure of the inner detector is clearly highlighted by over-densities of reconstructed vertices. These vertices are a large source of background and need to be removed. This is done using what is called a material map veto: using a spatial map of the materials

present in the inner detector², any vertex with coordinates (x, y, z) falling in a material region is completely removed. It is a strong requirement as it removes around 40% of the fiducial volume of the inner detector but is essential for analyses relying on secondary vertices. The effect of the material map veto can be observed in Fig. 7.6 in comparison with Fig. 7.5.

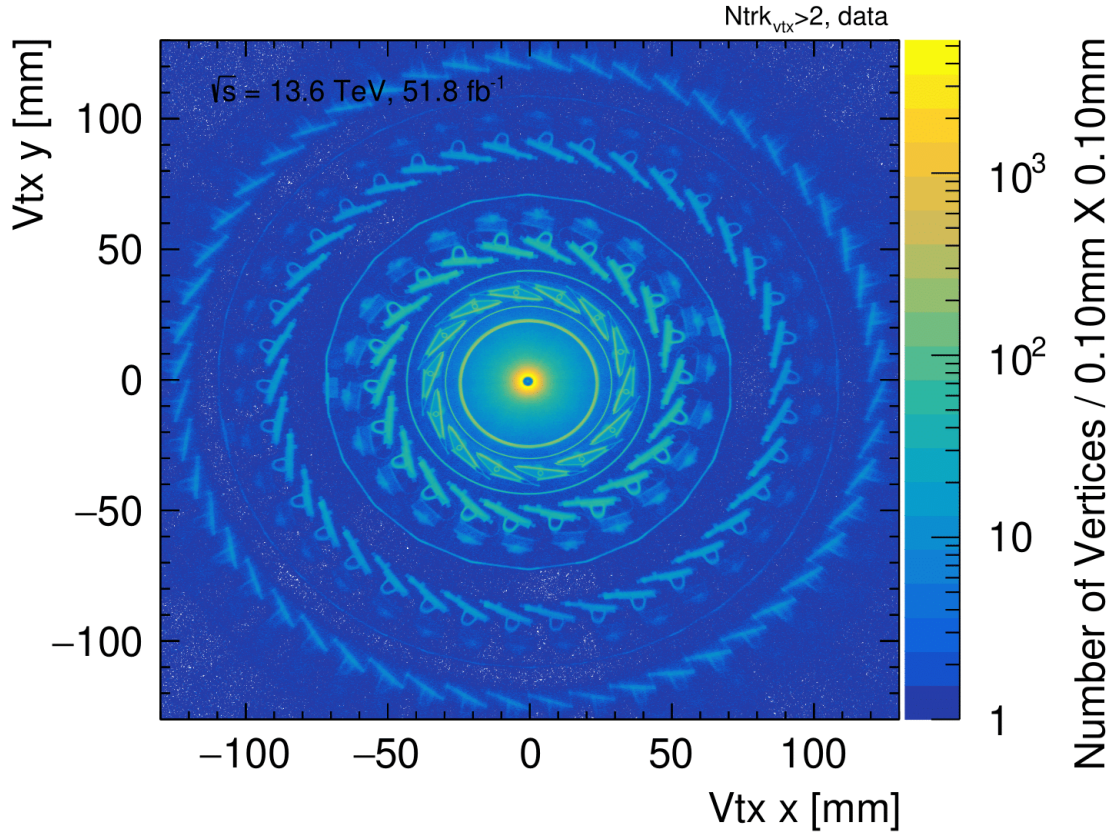


Figure 7.5: Two-dimensional histogram of the x and y coordinates of secondary vertices with $N_{\text{trk}} > 2$ in Run 3 data.

For three signal samples, the reconstruction efficiencies of the tracks and secondary vertices are shown in Fig. 7.7 as a function of the truth production radius (only a ratio with regard to one signal sample is shown for the vertex). While a precise determination of these efficiencies would require the use of low-level information on tracks (e.g. tracker hits), this is beyond the scope of the plots which are presented here for illustration purposes. Here, the tracking reconstruction efficiency is computed as the fraction of truth SM stable charged particles originating from the decay of a dark pion that can be geometrically (ΔR) matched to a reconstructed track. Similarly, the vertex reconstruction efficiency is computed as the fraction of truth dark pion particles decaying to at least two truth SM stable charged particles that can be geometrically ($\Delta = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$) matched to a reconstructed vertex. Both methods allow to provide a good enough approximation of the reconstruction efficiencies³ to obtain a sense of the behaviours and dependencies on the signal model parameters. The track reconstruction efficiency shows a very good efficiency close to the interaction point and then gradually decreases with increasing production radius. The behaviour of the track reconstruction efficiency is very close to the results obtain in the official study shown in Fig.

² This map is derived in data based on the positions of over-densities in the spatial coordinates of carefully selected vertices.

³ Compared to the official results previously presented in Fig. 5.5 and 5.6 in Sec. 5.2 and 5.3 respectively.

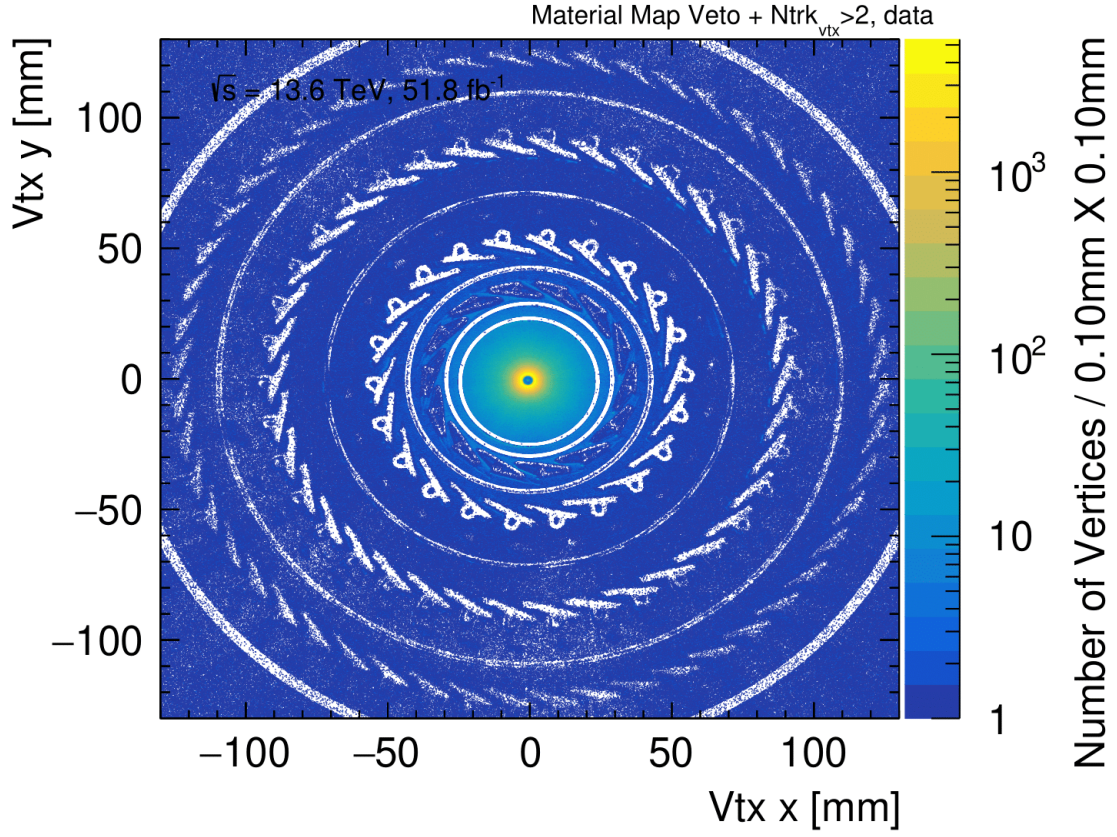


Figure 7.6: Two-dimensional histogram of the x and y coordinates of secondary vertices with $N_{\text{trk}} > 2$ after the material map veto in Run 3 data.

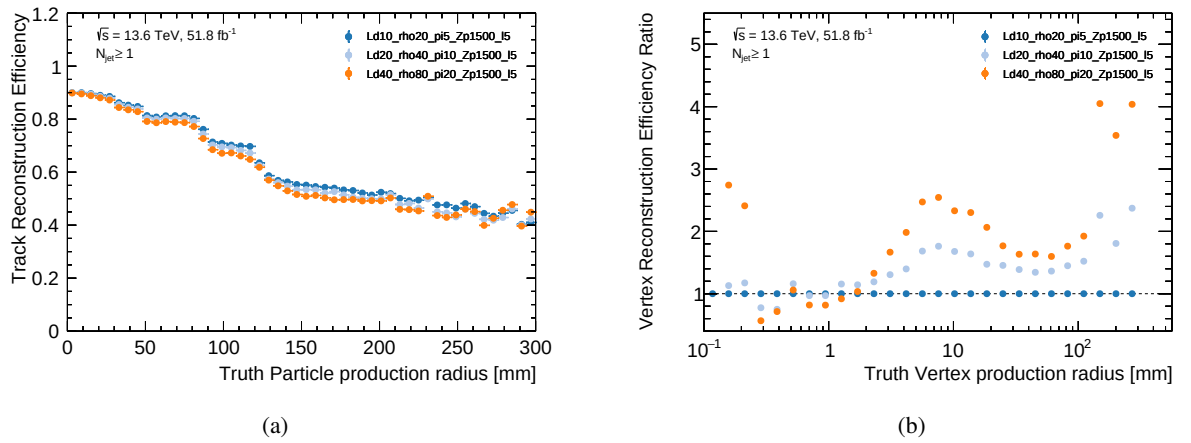


Figure 7.7: Track reconstruction efficiency (a) and vertex reconstruction efficiency ratio (b) for dark particle decays and products.

5.5. Almost no dependencies on the signal model parameters are observed. For the vertex reconstruction efficiency, only the ratio with regard to the signal sample `Ld10_rho20_pi5_Zp1500_l5` is shown to easily capture the dependence on the mass of the dark pion. The obtained vertex efficiency is thus observed to be much dependent on the mass of the dark pion with greater efficiency for higher mass. This can be expected as the greater the mass of the dark pion the more energy is provided to the pair of SM quarks and the more particles (and in particular charged particles) are produced per dark pion decay.

These two studies allows to globally understand the behaviours of the tracking and vertexing efficiencies in the signal samples. It mainly shows that the vertexing efficiency is dependent on the mass of the dark pion and that both the tracking and vertexing efficiencies have a strong connection with the production radius of the particles. Thus, the distribution of the production radii in the signal samples directly impacts the global reconstruction efficiencies of the dark pion vertices and the produced tracks, making the global reconstruction efficiencies dependent on the dark pion lifetime.

7.2.3 Special derivation: LLP1

This analysis uses a special derivation format called LLP1 which is commonly used by LLP analyses in ATLAS. This format is specialised for displaced object reconstruction. In order to be usable for this analysis, some additions were needed. I mainly participated in this development by adding: the reconstruction of the large- R RCEMTopo jet collection and the selections of tracks contained in these jets, the trigger needed (EJ and large- R jet triggers). I also developed an algorithm to compute substructure variables for the RCEMTopo jets based either on calorimeter clusters or tracks contained in the jets, these variables are used in the following background discrimination procedure. These additions led to a tolerable increase in disk space used by this derivation of 6.7% per event.

7.3 Trigger studies

As discussed in Section 4.3, in order to be recorded any data event needs to be selected by at least one trigger. In this analysis, three triggers are considered, two triggers based on jets and one single-muon trigger used to study the jet triggers and the agreement between data and MC simulations. Trigger efficiencies in data were measured to determine the regions of phase space where the triggers have a maximal, constant efficiency. The efficiencies in data are also compared to the efficiencies obtained in MC samples in order to verify that the trigger decisions are well modelled in MC simulations. The jet trigger efficiencies for MC signal samples are also presented briefly to confirm the relevance of the jet triggers for this analysis.

7.3.1 Muon trigger

The selection efficiency of a trigger is generally defined as the number of events selected by the trigger divided by the total number of events:

$$\text{Efficiency } \epsilon = \frac{\text{Number of events selected}}{\text{Total number of events}} \quad (7.2)$$

In data, calculating this efficiency is not straightforward as the total number of data events available is actually the number of events recorded by ATLAS and thus passing at least one trigger. It is not the total number of produced events and this bias the calculation of the efficiency. In order to actually compute this quantity it is necessary to use a secondary trigger independent of the first one to define a complete subset of data events and then use this subset to define the main trigger efficiency:

$$\text{Efficiency } \epsilon = \frac{\text{Number of events selected by both triggers}}{\text{Total number of events selected by the secondary trigger}} \quad (7.3)$$

In this analysis the triggers of interest that need to be studied are triggers based on jets. A single-muon trigger was considered in this analysis because the production of a muon in an event is supposed to be completely independent on the production of jets. A muon-based trigger is thus orthogonal to any jet triggers and can be used to compute the jet trigger efficiencies:

$$\text{Jet trigger efficiency } \epsilon = \frac{\text{Number of events selected by jet + muon triggers}}{\text{Total number of events selected by the muon trigger}} \quad (7.4)$$

Moreover, using a muon trigger allows to study jet observables in a control region (CR) without any pre-requirement on the jets as a second independent and uncorrelated object is used to select the events. Applying any requirements on muons will not bias the distributions of jet observables. This can be harnessed to study the agreement between MC simulations and data for jet observables and specifically the observables that are used in the analysis selections.

The chosen muon trigger is a standard trigger that select events with a reconstructed muon at the trigger level with $p_T > 24$ GeV. It is used to define a muon CR detailed in Table 7.3 in which the jet triggers and data/MC agreement are studied. This CR region is based, in addition to the muon trigger, on selections on a reconstructed tightly isolated muon. Requiring an isolated muon ensures that the muon used for the trigger decision was not produced inside a jet and is thus not correlated to the jet production. It also ensures that

no or few signal events should be selected in the muon CR as no isolated muon are considered in the signal phenomenology and only a muon production correlated to the jet dynamics can be expected.

Table 7.3: Muon control region selections

Single μ trigger	HLT_mu24_ivarmedium_L1MU14FCH
μ isolation	Tight
Number of μ	1
μ selection	$p_T > 27$ GeV
Number of jets	≥ 1
Jet selection	$p_T > 200$ GeV, $ \eta < 1.8$, pass OR with photon

7.3.2 Standard large- R jet trigger

The first jet trigger used in this analysis to select data events is a standard large- R jet trigger. It selects events containing at least one reconstructed large- R RCEMTopo jet at the trigger level with $p_T > 460$ GeV and no η requirement. It was chosen as it is based on the same jet collection as the one used in the analysis. The large- R trigger efficiency was measured in both data and MC background simulations (W +jets and top samples) as a function of the leading jet p_T following the procedure discussed above and is shown in Fig. 7.8. A fit of the efficiency in data was performed in order to determine the start of the plateau region where the efficiency is constant.

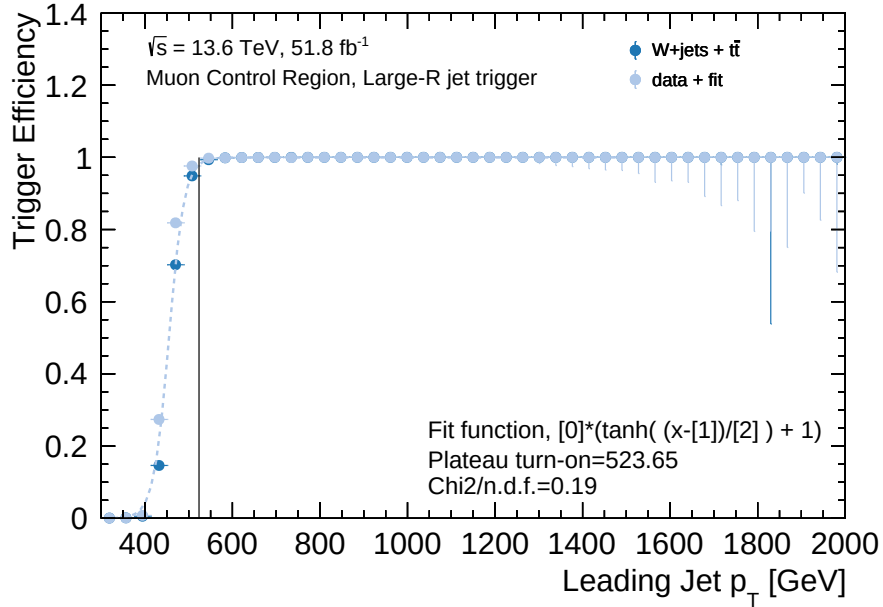


Figure 7.8: Large- R RCEMTopo jet trigger efficiency measured as a function of the leading jet p_T in data and MC in the muon CR. A fit of the efficiency in data is also shown. The fit in data is performed using the turn-on function $f(p_T, p, t, s) = p(\tanh((p_T - t)/s) + 1)$ where (p, t, s) are the fit parameters.

This trigger is shown to be fully efficient (i.e. $\epsilon > 98\%$) above a p_T of around 520 GeV and well modelled in MC simulations. The p_T threshold is higher than the initial trigger p_T requirement, the slight shift

is caused by the difference in energy between the jet reconstructed at the trigger level and the final jet reconstructed with the more precise offline software and which is used to compute the efficiency. These differences also create this turn-on effect between 400 GeV and the start of the plateau.

7.3.3 Run 3 Emerging Jet trigger

The second trigger used is an exotic jet trigger that was recently developed for the Run 3 called EJ trigger. This trigger aims at selecting events with an EJ-like signature by capitalising on the displaced object aspect of EJs. It selects events containing at least one reconstructed large- R PFlow jet with CSSK and SoftDrop procedures applied at the trigger level with $p_T > 200$ GeV, $|\eta| < 1.8$ and a prompt track fraction (PTF) below 0.08. The PTF measures the property of a jet to be composed of prompt tracks originating from the HS PV, it measures the jet p_T fraction originating from tracks with small d_0 displacements. It is defined as the p_T sum of tracks within a radius of $\Delta R = 1.2$ around a jet with $p_T > 1$ GeV, $\Delta z = |z_{\text{HS PV}} - z_0| < 10$ mm and $d_0 < 2.5\sigma_{d_0, \text{bkg}}(p_T^{\text{trk}})$ divided by the jet p_T :

$$PTF = \frac{\sum_{\text{trk} \subset \Delta R < 1.2} p_T^{\text{trk}}(d_0 < 2.5\sigma_{d_0, \text{bkg}}(p_T^{\text{trk}}))}{p_T} \quad (7.5)$$

$\sigma_{d_0, \text{bkg}}(p_T^{\text{trk}})$ is defined as the track d_0 significance, with $2.5\sigma_{d_0, \text{bkg}}$ used as a threshold to specify if a track has a d_0 significantly different from 0 and thus is not originating from the HS PV. It is a function of the track p_T and was derived in a background SM jet MC sample. For EJs, the PTF variable is small due to the numerous displaced tracks originating from secondary vertices and is thus a strong discriminant against SM QCD jets. The trigger requires a jet with $PTF < 0.08$.

As for the large- R jet trigger a study of the EJ trigger efficiency was performed in data and MC background samples. The obtained trigger efficiency as a function of leading jet p_T and PTF are shown in Fig. 7.9 and 7.10 respectively and a fit of the data is also presented. For the PTF observable it is not possible to exactly compute it with the same definition as in the trigger as only tracks within $\Delta R = 1.0$ around a jet are accessible, this is however a fairly similar proxy. To better observe the turn-on of the trigger in PTF the base-10 logarithm of PTF is used. Moreover, the jet p_T and PTF are two correlated variables, so, to avoid being impacted by this correlation, a cut is added on PTF in Fig. 7.9 and a cut on p_T in Fig. 7.10 to be in a region with constant efficiency in the opposite variable. Due to low statistics in the data and MC samples, the observation of a plateau is not evident but the fit of the efficiencies in data allows to define the two thresholds for the regions where the trigger is constant: $p_T \gtrsim 310$ GeV and $PTF \lesssim 0.05$. This trigger, by showing a p_T threshold lower than the standard large- R trigger, allows to be sensitive to events with a lower energy scale and in the case of the EJ events to be sensitive to lower Z' mass scenarios. In particular, this trigger largely increases the analysis sensitivity to the $m_{Z'} = 600$ GeV samples. However, discrepancies between data and MC were observed showing some mis-modelling of the trigger decision. These discrepancies were not well understood and led to difficulties for using this trigger. Thus, for this thesis it was favoured to focus on the standard large- R trigger and not consider the EJ trigger. It should however be used in the final ATLAS analysis.

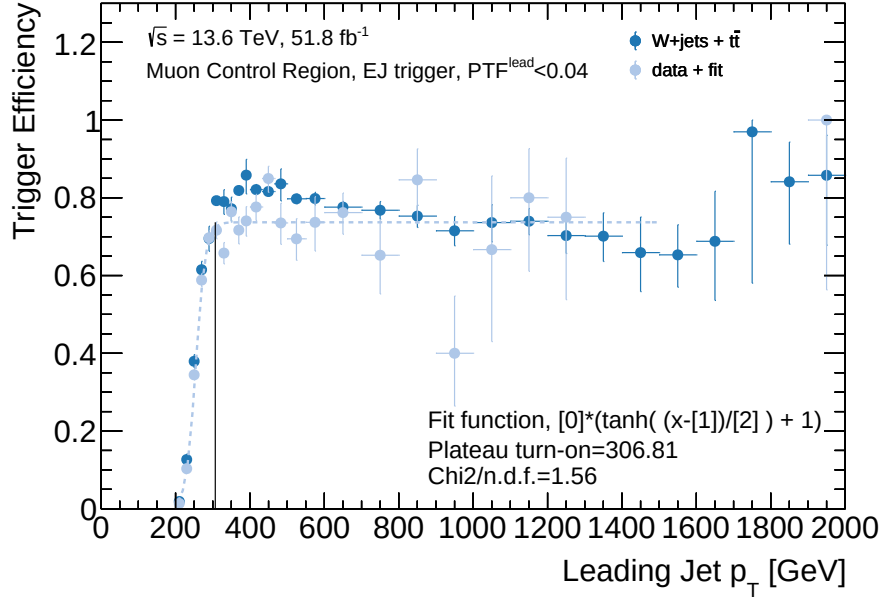


Figure 7.9: EJ trigger efficiency measured as a function of the leading jet p_T in data and MC in the muon CR with an additional selection on leading jet $PTF < 0.04$. The fit in data is performed using the turn-on function $f(p_T, p, t, s) = p(\tanh((p_T - t)/s) + 1)$ where (p, t, s) are the fit parameters.

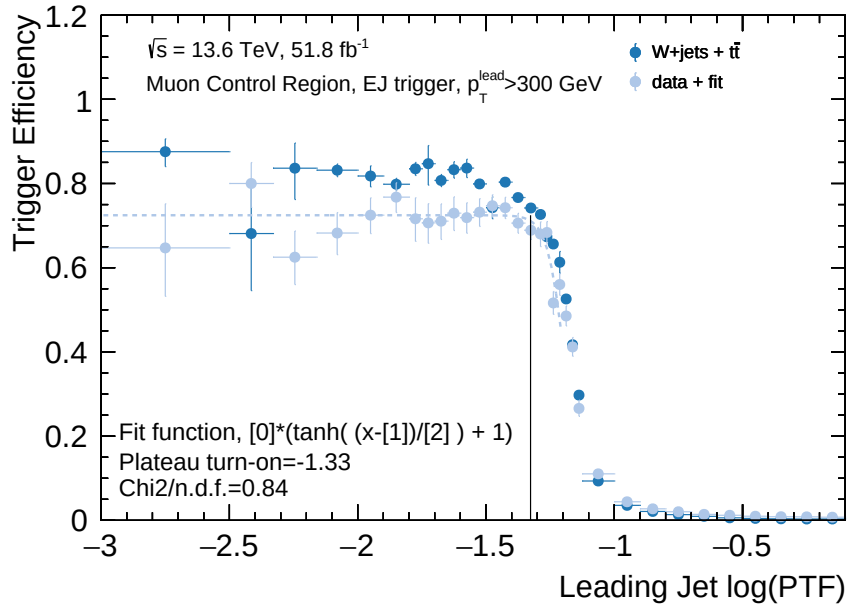
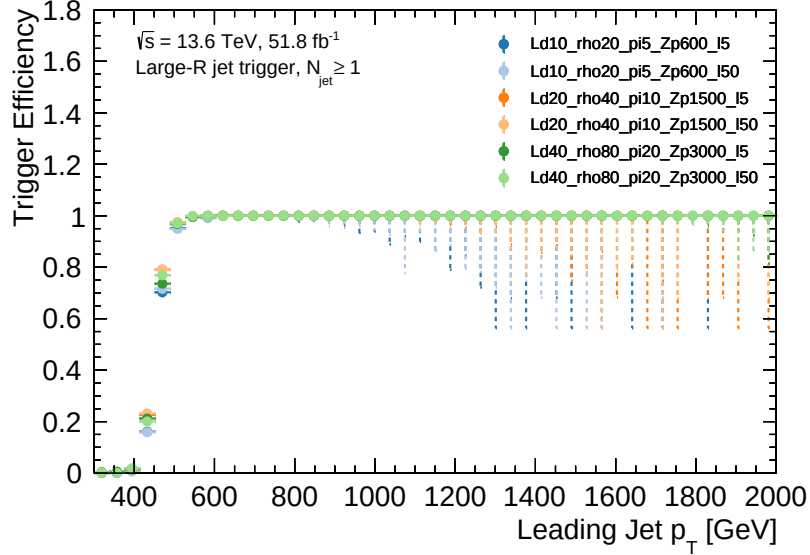


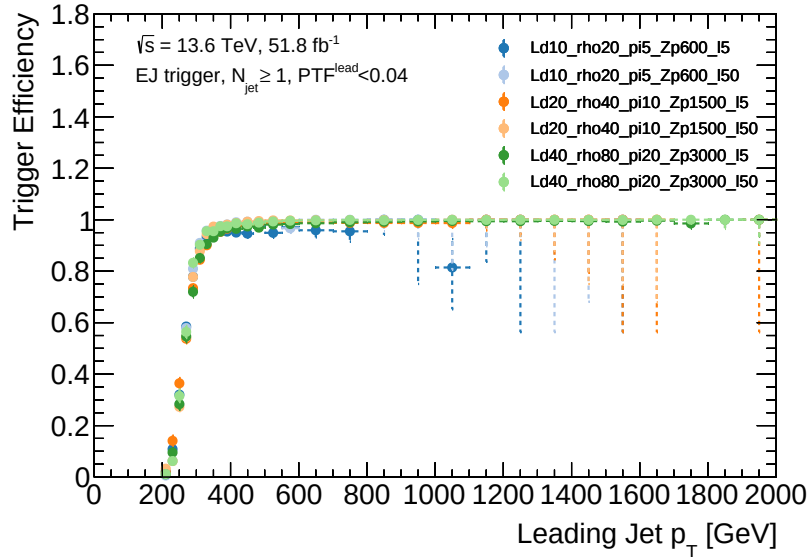
Figure 7.10: EJ trigger efficiency measured as a function of the leading jet $\log(PTF)$ in data and MC in the muon CR with an additional selection on leading jet $p_T > 300 \text{ GeV}$. The fit in data is performed using the turn-on function $f(p_T, p, t, s) = p(\tanh((p_T - t)/s) + 1)$ where (p, t, s) are the fit parameters.

7.3.4 Trigger efficiencies in MC signal samples

For the signal samples, to measure the trigger efficiencies it is not necessary to use the muon CR as MC samples are generated without any skimming by any triggers. The efficiencies can directly be measured using Eq. 7.2. Fig. 7.11 shows the trigger efficiencies in several MC signal samples. Both jet triggers are observed to be fully efficient in their respective plateau regions.



(a)



(b)

Figure 7.11: Large- R RCEMTopo jet trigger (a) and EJ trigger (b) efficiencies measured as a function of the leading jet p_T in MC signal samples in the muon CR with an additional selection on leading jet $PTF < 0.04$ in (b).

7.4 Analysis strategy

Searches for BSM physics are often searches for rare processes which are flooded in a widely dominant background contribution. To be sensitive to such rare processes it is necessary to define a region in which the background is greatly reduced and the signal is dominant. In such a region, generally referred to as the signal region (SR), one analysis strategy is then to count the number of events N_{evt} composing it and assess if there is an excess of events compared to a SM expected value of number of background events N_{bkg} :

$$N_{\text{evt}} > N_{\text{bkg}} \quad (7.6)$$

This is generally referred as a cut-and-count analysis and it is usually favoured for analyses that searches for exotic signature and for which it is rather easy to define a SR with high background reduction. This strategy revolves around two main aspects: the definition of a SR and thus the search for discriminating variables; and the ability of precisely predicting the SM background contribution in the SR.

The search for discriminant variables to evince EJs and the definition of the SR are presented in Sections 7.5 and 7.6. For the background estimation, a data-driven method is favoured. Discriminating a background from a highly exotic signature generally leads to define a SR in a very peculiar corner of phase space in which MC simulation mis-modelling can become a source of bias. To avoid uncertainties and biases in the background estimation in the SR, it is thus usual to use a method based solely on data. One of the most common data-driven background estimation methods is the ABCD method which consists in defining and using control regions (CRs), orthogonal to the SR in which the background is dominant, to infer the expected background contribution in the SR. This method is used in this analysis and will be detailed in this section.

As mentioned in Sec. 7.3, two triggers are considered in this analysis which are complementary in terms of the energy scale to which they are sensitive. The analysis is thus separated in two scenarios, one using the large- R jet trigger for high energy scale events (leading jet p_T above 520 GeV) and the second using the EJ trigger for lower energy scale. The two scenarios are separated based on a cut on the two-jet invariant mass m_{jj} at 1 TeV in order to have a clear separation between signal models with low Z' mass for which the EJ trigger is more sensitive and high Z' mass for which the large- R jet trigger is more suited. My work in this analysis was on the development of the large- R jet trigger based scenario and only this scenario is presented in this chapter.

7.4.1 ABCD data-driven background estimation method

The ABCD method requires that there are two selections on two observables, v_1 and v_2 , that form the region of phase space defined as the SR and which can be inverted to define three additional regions called CRs. The SR is supposed to be enriched in signal events and the CRs should be dominated by background events. These four regions form a two-dimensional plane called the ABCD plane as illustrated in Fig. 7.12.

In this plane the SR is usually associated to region A and the CRs to the regions B, C, D. For each region the number of events is the sum of the signal and background contributions:

$$N_{\text{evt}}^X = N_{\text{sig}}^X + N_{\text{bkg}}^X \quad (7.7)$$

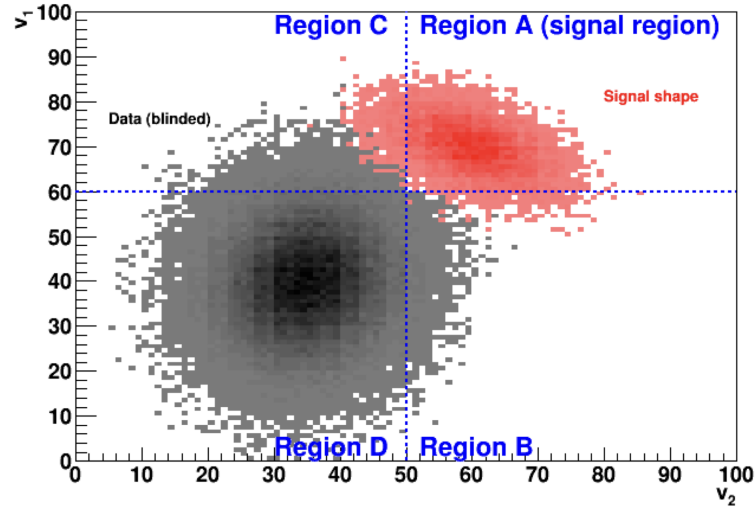


Figure 7.12: Illustration of an ABCD plane defined by selections on the observables v_1 and v_2 . The black distribution represents the data dominated by background events and the red distribution represents the hypothesised signal. Taken from Ref. [138].

where $X = (A, B, C, D)$. The ABCD method is then built on the assumption that for background events the following proportionality relation is true:

$$\frac{N_{\text{bkg}}^A}{N_{\text{bkg}}^B} = \frac{N_{\text{bkg}}^C}{N_{\text{bkg}}^D} \quad (7.8)$$

This assumption is verified if the two observables defining the ABCD plane are sufficiently uncorrelated for the background event distribution. If this is verified then an estimation of the background contribution in region A can be obtained by rearranging Eq. 7.8:

$$N_{\text{bkg,estimated}}^A = \frac{N_{\text{bkg}}^C}{N_{\text{bkg}}^D} N_{\text{bkg}}^B \quad (7.9)$$

Furthermore, as the CRs should be dominated by background events, the estimation can be approximated by:

$$N_{\text{bkg,estimated}}^A = \frac{N_{\text{evt}}^C}{N_{\text{evt}}^D} N_{\text{evt}}^B \quad (7.10)$$

Thus, by blinding the SR and using data events collected in the CRs it is possible to estimate the background contribution in the SR independently of information on the number of events in the SR. This is the basic principle of the ABCD background estimation method.

This method can be developed with the help of MC samples of background and signal to find the two defining observables and selections. A validation of the method can be done by defining a validation region

(VR) A' and new CRs B', C', D' in a subpart of the ABCD plane which exclude the SR. This is illustrated in Fig. 7.13. The estimation method can then be used with these new CRs:

$$N_{\text{bkg,estimated}}^{A'} = \frac{N_{\text{evt}}^{C'}}{N_{\text{evt}}^{D'}} N_{\text{evt}}^{B'} \quad (7.11)$$

The new selections defining A', B', C' and D' can be varied to ensure that the two observables are well-uncorrelated in background events in most of the ABCD plane. This validation technique is useful to validate the ABCD method in MC, but most importantly in data without having to unblind the SR.

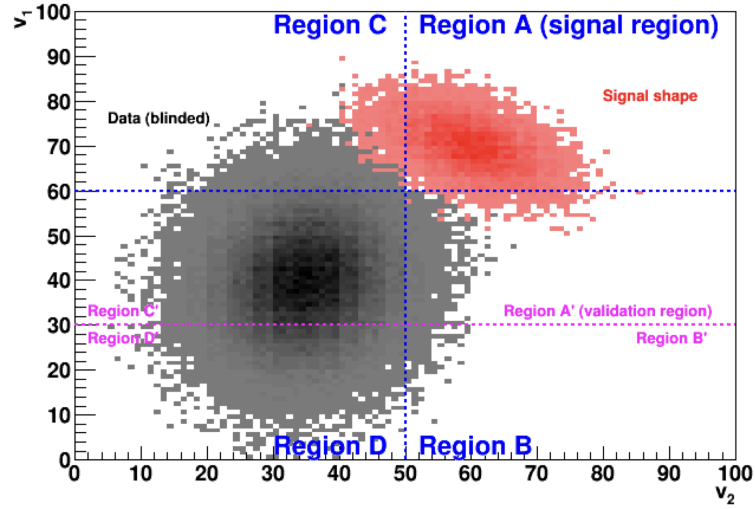


Figure 7.13: Illustration of a sub-ABCD plane for validation of the method. Taken from Ref. [138].

In order to be more consistent and to take into account possible signal contamination in the CRs, it is also possible to take a likelihood-based approach to the ABCD method. It consists in fitting a statistical model constructed as a four-bin model, each bin corresponding to the number of observed events in one of the four regions of the ABCD plane, and enforce the ABCD relation on the expected number of background events in each region. In this model, the expected number of events per region $N_{\text{evt}}^X = N_{\text{sig}}^X + N_{\text{bkg}}^X$ can be parameterised by four free parameters $(\mu, \mu_B, \tau_B, \tau_C)$ such that:

$$\begin{aligned} N_{\text{evt}}^A &= \mu + \mu_B \\ N_{\text{evt}}^B &= \mu q_B + \mu_B \tau_B \\ N_{\text{evt}}^C &= \mu q_C + \mu_B \tau_C \\ N_{\text{evt}}^D &= \mu q_D + \mu_B \tau_B \tau_C \end{aligned}$$

μ is the number of signal events in the SR (later referred to as signal strength). μ_B is the number of background events in the SR and τ_B and τ_C are the parameters enforcing the ABCD relation on the background distribution:

$$\frac{N_{\text{bkg}}^A}{N_{\text{bkg}}^B} = \frac{\mu_B}{\mu_B \tau_B} \equiv \frac{N_{\text{bkg}}^C}{N_{\text{bkg}}^D} = \frac{\mu_B \tau_C}{\mu_B \tau_B \tau_C}$$

The q_X are not free parameters and only represent the signal leakages, $N_{\text{sig}}^X/N_{\text{sig}}^A$, in the CRs B, C, D. The likelihood for observing the actual number of data events in each region (**data** = $\{N_{\text{data}}^A, N_{\text{data}}^B, N_{\text{data}}^C, N_{\text{data}}^D\}$) can then be constructed as the product of four Poisson distributions, each corresponding to one region, with fixed observation N_{data}^X and adjustable expectation N_{evt}^X :

$$\begin{aligned} \mathcal{L}(\mathbf{data}|\mu, \mu_B, \tau_B, \tau_C) = & \text{Pois}(N_{\text{data}}^A | \mu + \mu_B) \\ & \times \text{Pois}(N_{\text{data}}^B | \mu q_B + \mu_B \tau_B) \\ & \times \text{Pois}(N_{\text{data}}^C | \mu q_B + \mu_B \tau_C) \\ & \times \text{Pois}(N_{\text{data}}^D | \mu q_D + \mu_B \tau_B \tau_C) \end{aligned}$$

This likelihood can then be maximised to obtain the values of the free parameters characterising the signal and background that are the most suited to the observed data but also be used to test different signal or background-only hypotheses. The likelihood approach of the ABCD method is used for the final statistical analysis (and will be completed in Sec. 7.9), while the standard approach is used for simplicity in the development of the method using MC samples.

7.5 Discriminating observables and data-MC agreement

In the analysis strategy, the first stage is to study, in MC simulations, the signal characteristics and their potential dependence on the signal parameters in order to find discriminating observables against the background. My contribution was to develop the analysis scenario based on the large- R jet trigger, thus the signal studies were made on events selected by this trigger. Additional event pre-selections, forming the baseline selection for the discriminating studies, are also considered and are summarised in Table 7.4. At least two jets are required with the leading jet p_T above the trigger plateau threshold of 520 GeV and the sub-leading jet p_T is required to be above 300 GeV. This is to ensure that the two leading jets form a dijet pair with balanced p_T . The two leading jets must have $|\eta| < 1.5$ so that all the jets areas are within the tracker acceptance ($|\eta| < 2.5$). As in the muon CR, the two leading jets must not overlap with an isolated photon.

Table 7.4: Baseline selection for discrimination studies.

Large- R jet trigger	HLT_j460_a10r_L1J100
Number of jets	≥ 2
Leading Jet	$p_T > 520$ GeV, $ \eta < 1.5$, pass OR with photon
Sub-leading Jet	$p_T > 300$ GeV, $ \eta < 1.5$, pass OR with photon

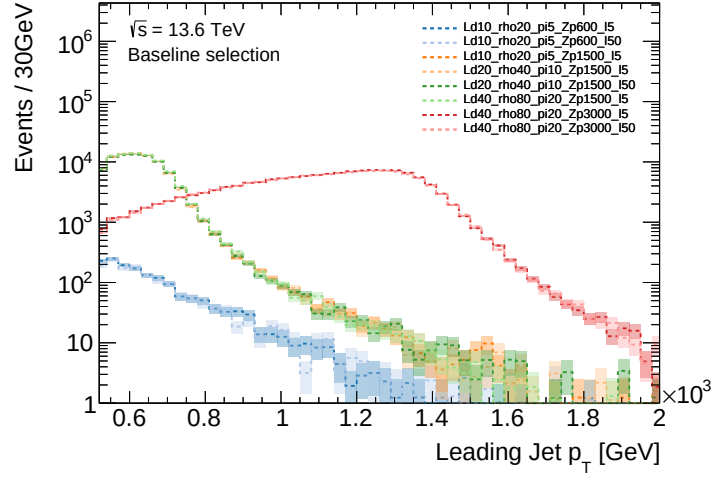
This section will first present basic signal kinematics and the expected specificities of the signal. Two subsections will follow presenting the discrimination between signal and background for observables based on displaced objects and based on the jet substructure. In addition to the signal and background discrimination studies, the agreement between data and MC simulations in the muon CR (introduced in Sec. 7.3) for the observables of interest will be discussed. The agreement between data and simulation is crucial for two aspects. The analysis development is based on MC simulations, it is thus necessary to ensure that the MC simulations are well representative of the data even though the final estimations are entirely data driven. Moreover, the inferred conclusion on the EJ signal are based on the MC signal sample yields and acceptances in the SR.

7.5.1 Signal characteristics

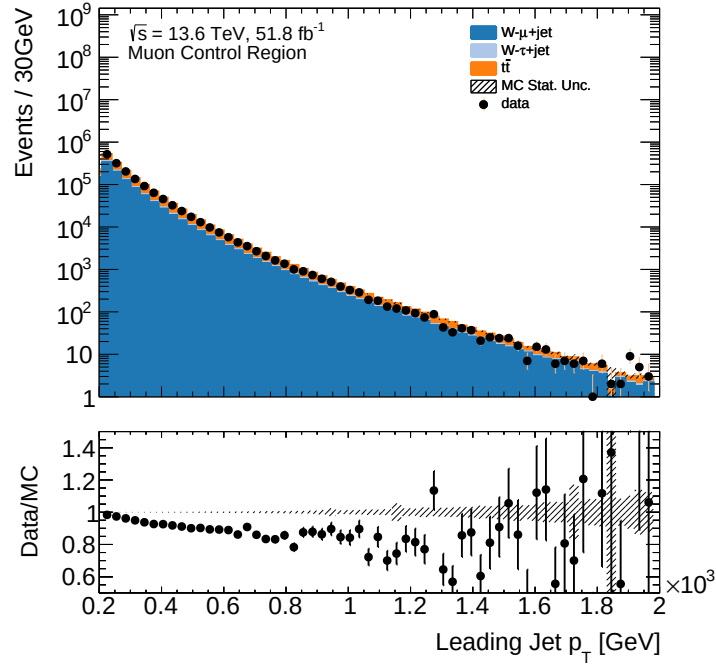
The signal basic kinematics (leading jet p_T , m_{jj} and leading jet η) distributions are shown in Fig. 7.14, 7.15 and 7.16 along the corresponding data/MC agreement plots in the muon CR. The muon CR is used here (and in the rest of the section) to blind the data to any signal contamination (no isolated muons are expected in the EJ signals). For the signal, the production cross-section is a priori not known and depends on the arbitrary choice of coupling of the Z' to SM quarks and the mass of the Z' . The figures thus simply show the number of generated events passing the baseline selection, where the total number of generated events equals 200,000 for each signal sample.

The p_T and m_{jj} distributions (Fig. 7.14 and 7.15) show as expected a strong correlation with $m_{Z'}$. For the models with $m_{Z'} = 600$ GeV, the pre-selection on the two leading jets p_T kills the acceptance which makes the large- R jet trigger analysis mostly insensitive to these models. The focus of this analysis thus will be on the higher $m_{Z'}$ models (1.5 and 3 TeV). In the following discriminating studies, the distributions for the models with $m_{Z'} = 600$ GeV will be shown for indication but will not be considered for the optimisation of the analysis selections. The two-jet invariant mass distributions in signals also show that these p_T selections select events with an invariant mass mostly above 1 TeV which motivates the separation on m_{jj}

between the two analyses. Very small dependence is observed on the dark pion lifetime and mass. Both observables are not very well modelled in MC with data/MC ratios presenting clear decreasing slopes. This discrepancy is assumed to come from the known mis-modelling of the jet p_T in $t\bar{t}$ MC samples [139]. It is not expected to be problematic for the analysis as the $t\bar{t}$ events are a largely sub-dominant background. Moreover, the modelling of the jet p_T and m_{jj} is found to be better in an analysis-like phase-space region and in the dominant MC background dijet sample. This will be presented and discussed in Sec. 7.7.

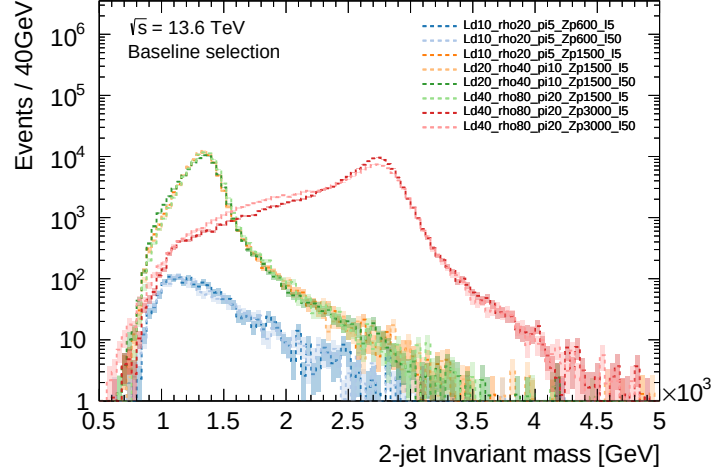


(a)

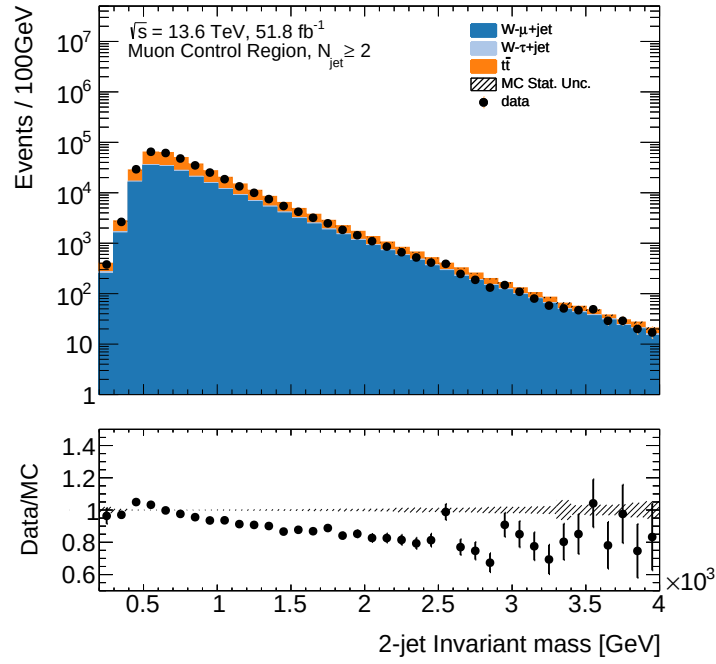


(b)

Figure 7.14: Leading jet p_T in several MC signal samples after the baseline selection defined in Tab. 7.4 (a) and leading jet p_T in data and MC W +jet and top samples in the muon CR (b).



(a)



(b)

Figure 7.15: 2-jet invariant mass in several MC signal samples after the baseline selection defined in Tab. 7.4 (a) and leading jet η in data and MC W +jet and top samples in the muon CR with $N_{jet} \geq 2$ (b).

The distribution of the leading jet η (Fig. 7.16) shows no particularities nor dependences on the signal parameters. It is well modelled in MC with a flat data/MC ratio close to one. The observations on the leading jet p_T and η are similar to the ones on the sub-leading jet.

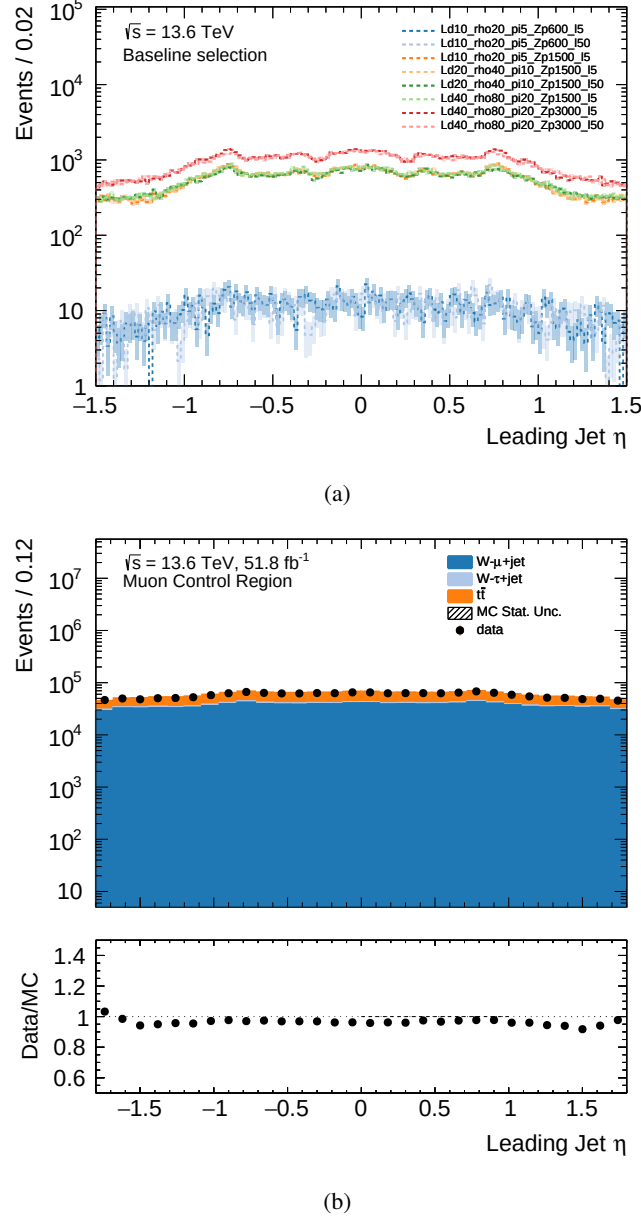


Figure 7.16: Leading jet η in several MC signal samples after the baseline selection defined in Tab. 7.4 (a) and leading jet η in data and MC $W+\text{jet}$ and $t\bar{t}$ samples in the muon CR (b).

The specificities of the signal compared to the background is expected to revolve around the displaced object aspect of the EJs but also in their particular internal structure arising from the dark QCD sector dynamic. Observables based on tracks and vertices are good candidates to harness the displaced object properties. Particularities in the jet internal structure can be measured through standard jet substructure observables that are used in ATLAS to tag jets that are not directly produced from SM QCD processes.

These observables are generally used for large- R jets to differentiate SM QCD jets from hadronic decay of boosted massive particles like Higgs and W/Z boson or top quark [105–109], by trying to identify the multipronged structure of these jets (i.e. composed of subjets).

For all the following discrimination studies, the MC background is re-weighted as discussed in Section 7.1 but not the signal for the reason discussed above. In order to be able to compare the observable distributions between background and signal without being biased by the arbitrary signal yield, all histograms are normalised by their integral such that only the shapes of the distributions are compared. In addition, ratios between signal and background are presented to ease the interpretation.

7.5.2 Displaced objects

The first obvious specificity of EJs is the presence of reconstructed secondary vertices within the EJs. Secondary vertices are rarely expected in the dominant dijet background events. They are, however, more common in the sub-dominant $t\bar{t}$ background. Vertices can be produced in the case where jets are originating from a bottom quark or contain a bottom quark from the decay of a top quark ($t \rightarrow q\bar{q}b$ or $t \rightarrow l\nu_l b$), as resulting B mesons can be long-lived and decay after travelling some distance in the tracker. However, the b-jets or top jets are not the dominant background. Background vertices, independently of the background production process, can also come from the decays of kaons and interactions with the tracker material. Selections on the vertices (mass, number of tracks and material map veto) as discussed in Sec. 7.2.2 allow to reduce these sources of background vertices. The number of selected secondary vertices within a jet $N_{\text{vtx,sel}}$ is thus a good candidate for discriminating signal from background. Fig. 7.17 shows $N_{\text{vtx,sel}}$ for the leading jet in background and several signal samples along the corresponding data/MC agreement plot in the muon CR.

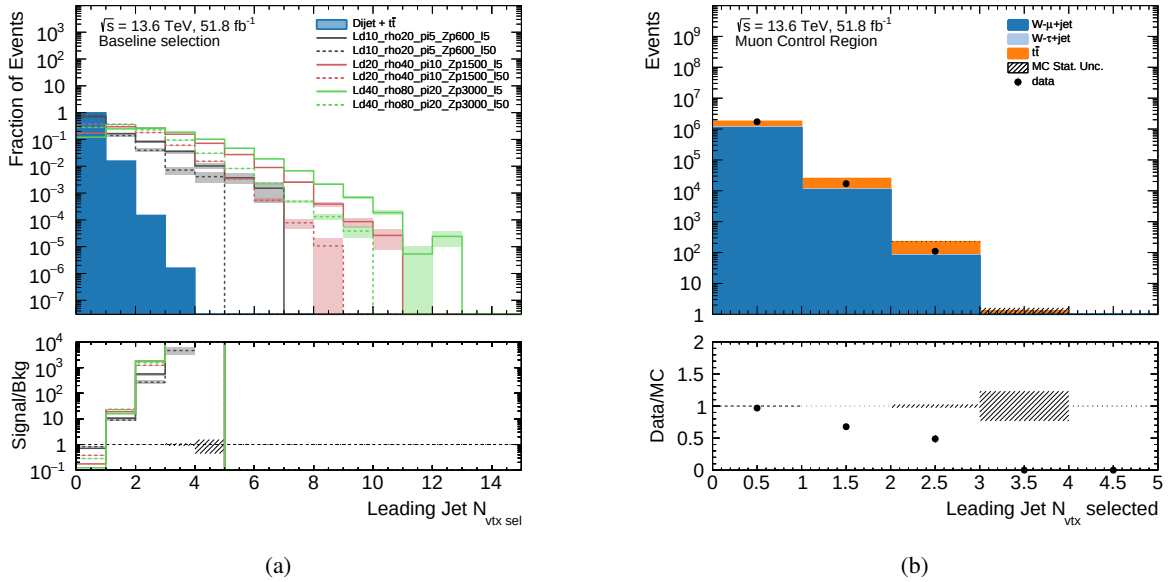


Figure 7.17: Leading jet number of selected secondary vertices in MC background and signal samples after the baseline selection defined in Tab. 7.4 (a) and leading jet number of selected secondary vertices in data and MC W +jet and top samples in the muon CR (b).

It shows a very strong discrimination power: selecting on the leading jet $N_{\text{vtx,sel}}$ to be above or equal to one removes around 99% of the background events while preserving a large fraction of the signal⁴. Similar discriminating power is observed for the sub-leading jet.

$N_{\text{vtx,sel}}$ seems, however, to be poorly modelled by the MC simulations with around 30% less data for events with at least one vertex and above 50% for higher $N_{\text{vtx,sel}}$. This discrepancy is mainly originating from the application of the material map veto on the secondary vertices. The material map veto is derived using data events and is thus representative of the actual position of the tracker materials within the detector, it is thereby correctly configured to remove material interactions in data events. However, it is not the case for MC simulations as the detector configuration and in particular the position of the tracker supplied to the detector simulation is not exactly mimicking the real-life detector. This leads to discrepancies in the removal of vertices with the material map veto in MC simulations. This effect is visible when constructing a plot of the x and y coordinates of vertices with $N_{\text{trk}} > 2$ after the material map veto in MC simulations. Fig. 7.18 shows this 2D histograms for the dominant MC dijet background with a focus on the centre of the tracker, remaining over-densities are clearly visible at radii around 20 mm which are not observed in data as shown in Fig. 7.6. These over-densities corresponds to the beam pipe not being perfectly covered by the data-based material map veto. It is expected that this will not impact much the analysis for two reasons: first, the background estimation is data-driven and the discriminating power of $N_{\text{vtx,sel}}$ should be even greater in data as there should be less material interaction vertices left in data; second, the measurements of the signal acceptance in the SR should not be impacted by this as, in signal events, most of the reconstructed vertices are actually coming from the real decay of particles and the total fiducial volume removed by the material map veto is identical between data and MCs.

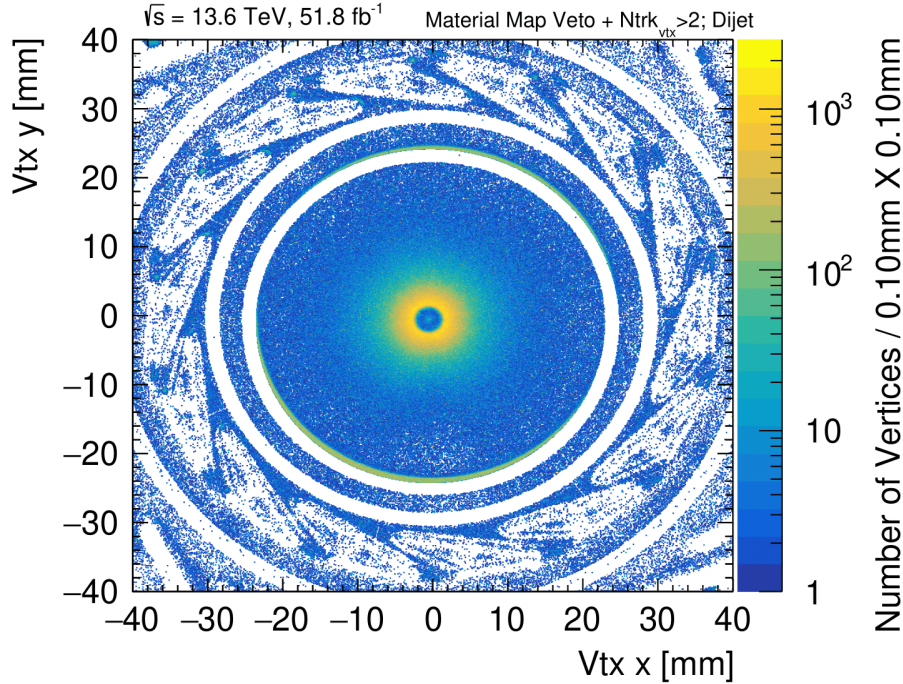


Figure 7.18: Two-dimensional histogram of the x and y coordinates of secondary vertices with $N_{\text{trk}} > 2$ after the material map veto in dominant MC dijet background (no event selection).

⁴ Except for the lowest $m_{Z'}$ models, but as discussed these models are not considered for the optimisation of the selections.

Selections on the leading and sub-leading jet $N_{\text{vtx,sel}}$ above or equal to one are the first considered discriminating selections to use for defining the SR.

The second tracking specificity of the EJs, which is correlated to the secondary vertices, is the presence of tracks not originating from the HS PV, produced by the decays of the dark pions. This is again not standard for SM QCD jets. This aspect is already well harnessed by the EJ trigger and the PTF observable. PTF is thus a priori a very good candidate to discriminate signal from background. The definition of PTF is reminded here:

$$PTF = \frac{\sum_{\text{trk} \in \Delta R < 1.0} p_T^{\text{trk}}(d_0 < 2.5\sigma_{d_0, \text{bkg}}(p_T^{\text{trk}}))}{p_T} \quad (7.12)$$

where the sum runs over the tracks found within $\Delta R = 1.0$ around the jet with $p_T > 1$ GeV and $\Delta z = |z_{\text{HS PV}} - z_0| < 10$ mm. Fig. 7.19 shows the normalised distributions of the leading jet PTF for background and signal along the corresponding data/MC agreement plot. PTF shows a very good discriminating power with the majority of the signal events close to $PTF = 0$ and the background distribution spreading over all the PTF range and falling for low values of PTF . A clear separation between signal and background for $PTF < 0.2$ is observed for all models. A peculiar behaviour is observed for the models with $m_{Z'} = 600$ GeV, which distributions seem to mimic the background distribution above $PTF = 0.2$. It is supposed that, for these models, the strong requirement on the leading jet p_T tends to select events where the leading jet is not an EJ but instead a SM pileup jet. A similar observation is made for the sub-leading jet PTF . PTF is also shown to be rather well modelled in MC simulation with a mostly flat data/MC ratio within 10% of 1.

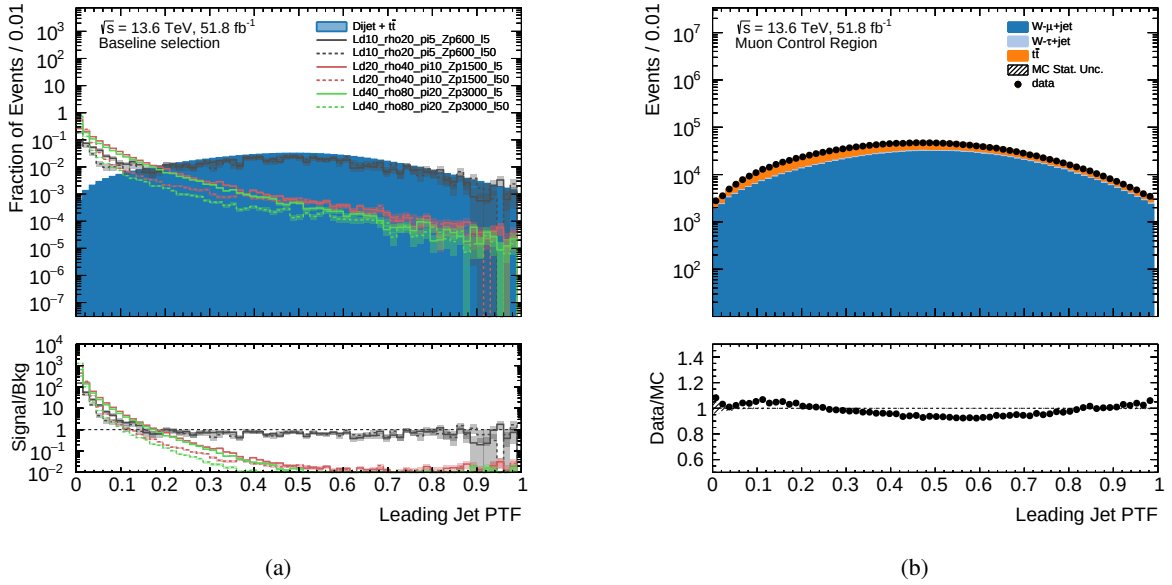


Figure 7.19: Leading jet PTF in MC background and signal samples after the baseline selection defined in Tab. 7.4 (a) and leading jet PTF in data and MC W +jet and top samples in the muon CR (b).

However, a bias in the PTF observable arises from its definition, from the track selection. Tracks entering the calculation of PTF are required to have $\Delta z = |z_{\text{HS PV}} - z_0| < 10$ mm which supposes that the reconstructed PV selected as the HS (i.e. the one with the largest squared sum of the p_T of the associated

tracks) is the one that produced the two leading jets. If this is not the case (if the selected PV is a pileup vertex) the computation of PTF can be largely biased as the selection on Δz will possibly discard tracks that are within the leading jets. This can lead to a lower computed PTF or even $PTF = 0$ if the wrongly selected PV is at a distance greater than 10 mm of the actual HS vertex. This wrong selection is expected to be very rare for background events as the leading jets are supposed to have many tracks pointing to the HS PV but there are very peculiar cases where it can happen and which lead to events with both leading jets with a very low PTF . For signal, it can be more frequent as there are many tracks within an EJ that are not originating from a PV, it is however not problematic as PTF is expected to be low in any case.

Fig. 7.20 shows, for background events, the distribution of the difference in the z coordinate between the selected reconstructed PV and the true generated HS PV. It shows as expected that the vast majority of events select the right PV but, even though rare, as the dijet statistic is very high, there is a non-negligible number of events for which the wrong PV is selected. This was observed to be a potential source of bias for the analysis, especially in the background estimation, as a residual number of background events can survive the discriminating selections on PTF .

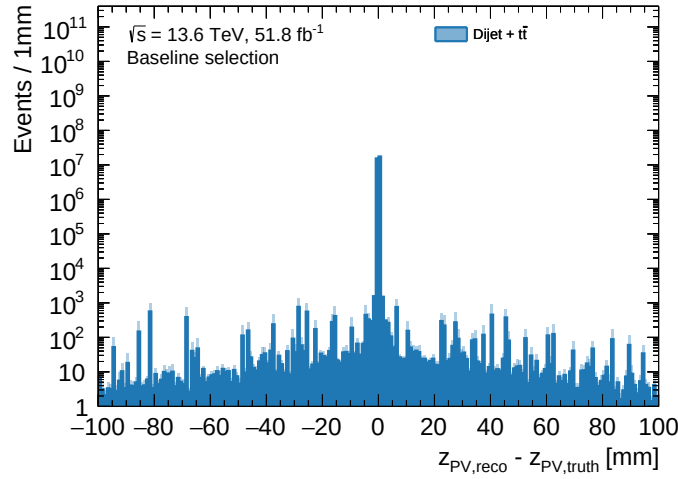


Figure 7.20: Difference in the z coordinate between the selected reconstructed PV and the true generated HS PV in MC background after the baseline selection defined in Tab. 7.4.

To remedy such a bias, a first solution is to remove the selection on Δz for the tracks entering the calculation of PTF . The resulting normalised distributions for signal and background are shown in Fig. 7.21 along the data/MC agreement plot. Similar observations as for the nominal PTF are made for both the discrimination power of this new observable and its rather good MC modelling. However, the selection on Δz allowed to limit the contamination from pileup tracks in the computation of PTF . The pileup contamination can be clearly seen in the signal distributions which do not peak any more at $PTF = 0$. This can be a problem as the PTF variable can now be quite dependent on the level of pileup, this solution was thus not favoured.

The second examined solution is to combine the two definitions of PTF , the nominal one and the no Δz track selection one, to construct an additional observable taken as the absolute difference between the two defined PTF and referred to as ΔPTF . Events with the right selection of HS PV should not be strongly affected by the change in PTF definition and should lead to small ΔPTF . However, events selecting the wrong HS PV should be much more affected by this change in definition and lead to larger ΔPTF . Thus selecting out events with large ΔPTF is expected to remove events with a biased measurement of PTF .

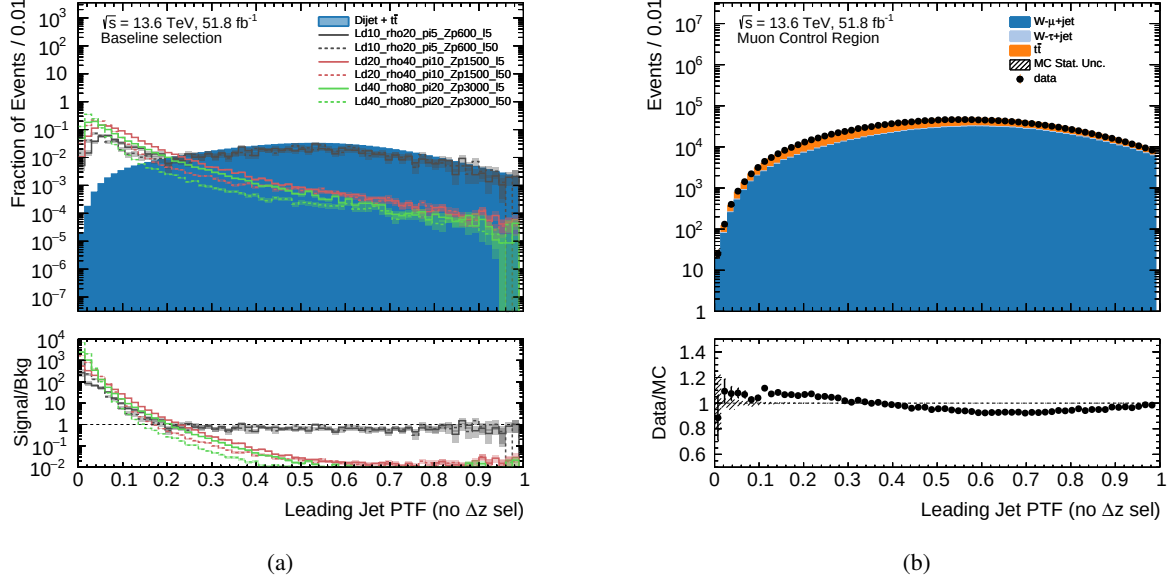


Figure 7.21: Leading jet PTF without Δz track selection in MC background and signal samples after the baseline selection defined in Tab. 7.4 (a) and leading jet PTF without Δz track selection in data and MC W +jet and top samples in the muon CR (b).

Fig. 7.22 shows the normalised distribution of ΔPTF for signals and background along the usual data/MC agreement plot. Signal and background have a similar distribution with most of the events concentrated at low values of ΔPTF corresponding to events with the right selection of HS PV or at least with a selected PV not far from the actual HS vertex. The MC modelling for this observable is not as good as for the two PTF observables but the agreement between data and MC is best in the bulk of the distribution such that the mis-modelling is not concerning. Downstream of the analysis, as will be shown later, it was found that a selection on the leading and sub-leading jet ΔPTF below 0.4 allowed to remove residual background events in the SR. The effect of this selection can already be observed in the distribution of the difference in the z coordinate between the selected reconstructed PV and the true generated HS PV as shown in Fig. 7.23. A clear reduction in the number of events at high values is observed⁵ in comparison to Fig. 7.20.

From the displaced object aspect of EJs two discriminating variables are considered, with four corresponding potential selections defining the SR:

- leading and/or sub-leading jet $N_{\text{vtx,sel}} \geq 1$,
- leading and/or sub-leading jet nominal $PTF < 0.2$.

In addition, a cleaning selection is also considered: $\Delta PTF < 0.4$ for the leading and sub-leading jet.

⁵ The high remaining peaks are composed of only one event with a very large MC weight w_{MC} . These kinds of events are rare but expected in the MC dijet sample. This will be discussed further in Sec. 7.7.

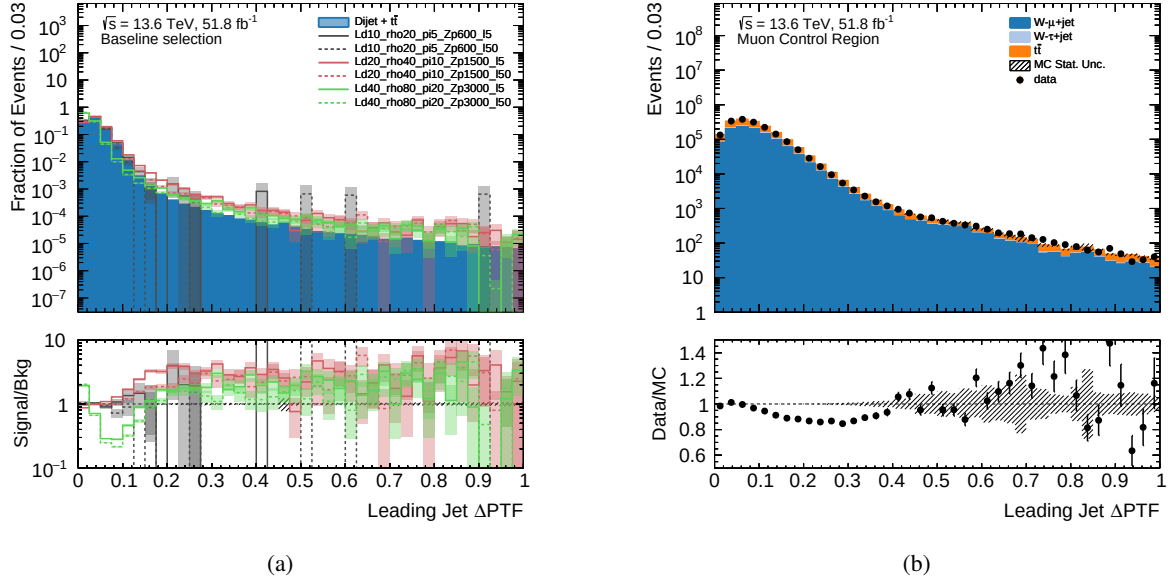


Figure 7.22: Leading jet ΔPTF in MC background and signal samples after the baseline selection defined in Tab. 7.4 (a) and leading jet ΔPTF in data and MC W +jet and top samples in the muon CR (b).

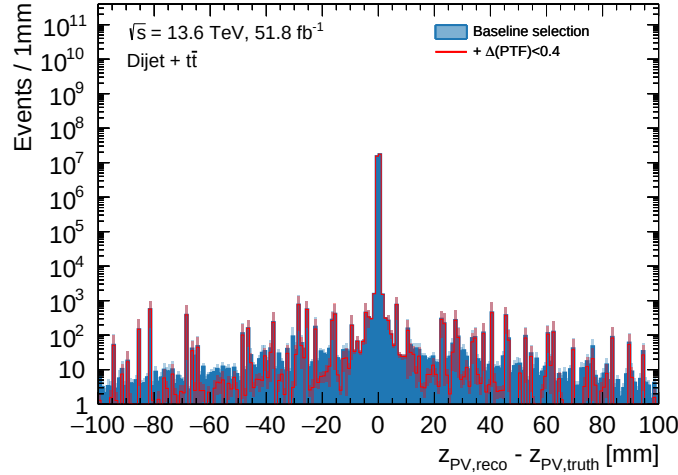


Figure 7.23: Difference in the z coordinate between the selected reconstructed PV and the true generated HS PV in MC background after the baseline selection defined in Tab. 7.4 with (red) and without (blue) the selections on the leading and sub-leading jet $\Delta PTF < 0.4$.

7.5.3 Jet substructure

The second specificity of EJs that can be capitalised on is their particular internal structure. Multiple substructure observables are used in ATLAS to distinguish SM QCD jets (one pronged) and boosted massive particle decays (multipronged). These observables can be repurposed to differentiate EJs from SM QCD jets. EJs are expected to have a multipronged-like internal structure as each dark pion decay produces a separate hadronic shower. Four different substructure measurement techniques were considered, each corresponding to a different way to characterise the internal structure of a jet.

The first consideration is the N-point energy correlation functions [106] for which the most promising observable for discriminating EJs from SM QCD jets was found to be the 2-point energy correlation function $ECF2$ defined by:

$$ECF2 = \sum_{i < j \in J} p_T^i p_T^j \Delta R_{ij} \quad (7.13)$$

where the sum runs over the jet (J) constituents. This observable measures the property of a jet to be composed of two or more internal subjet-like structures. $ECF2$ is expected to be small if a jet is composed of less than two subjets and vice versa. The N-subjettiness [108] can also be used and the 2-subjettiness τ_2 was found to be the best discriminating observable for this technique. It is computed by re-clustering the jet constituents with the k_t jet clustering algorithm and requiring that exactly two subjets are found. The two subjets are then used to define τ_2 :

$$\tau_2 = \frac{1}{d_0} \sum_k p_T^k \min(\Delta R_{1,k}, \Delta R_{2,k})$$

$$d_0 = \sum_k p_T^k R$$

where k runs over the jet constituents, $\Delta R_{j,k}$ is the angular distance between the constituent k and the subjet j and R is the jet radius used in the original jet clustering (here $R = 1.0$). This observable measures the property of a jet to be well-defined by a two-pronged internal structure, meaning that all constituents of the jet are localised near one of the two subjets. A small value of τ_2 thus corresponds to a jet looking like a two-pronged object. The third reviewed substructure technique was the k_t -splitting scale [105] obtained by re-clustering the jet constituents with the k_t clustering algorithm. The splitting scale is obtained by stopping the clustering at the step $N - i$ (where N is the total number of steps) and taking the k_t -distance between the two proto-jets that would be clustered at the step $N - i + 1$. The most interesting observable was found to be the splitting scale obtained at the step $N - 2$ (three remaining proto-jets), referred as *Split23* and defined by:

$$Split23 = \sqrt{d_{23}} = \min(p_T^{j2}, p_T^{j3}) \Delta R_{j2,j3} \quad (7.14)$$

where $j2$ and $j3$ are the two proto-jets to be merged at the clustering step $N - 1$. A large value of *Split23* thus indicates that a jet is composed of more than two subjets. Finally, the observable Q_w was also considered. It is defined as the smallest invariant mass among the proto-jet pairs of the last 3 steps of the k_t re-clustering sequence applied to the jet constituents. A large value corresponding to a jet with a large angular distance between two subjets.

These four observables are first computed using the topo-clusters within a large- R RCEMTopo jet (same constituent type as used in the definition of the jets) in order to have observables completely independent of the tracks that are already used to compute the previous discriminating variables. As all topo-clusters within a jet are used, the SoftDrop grooming technique is applied to reduce the pileup contamination in the computations of the substructure observables, in agreement to what is done for the reconstruction of UFO large- R jets. In addition, $ECF2$, $Split23$ and Q_W are not dimensionless with $ECF2$ in GeV^2 , and $Split23$ and Q_W in GeV , these three observables are thus divided by the p_T of the jet to limit the dependency on the energy scale of the event. Fig. 7.24 presents the normalised distributions of the leading jet substructure observables in signal and background.

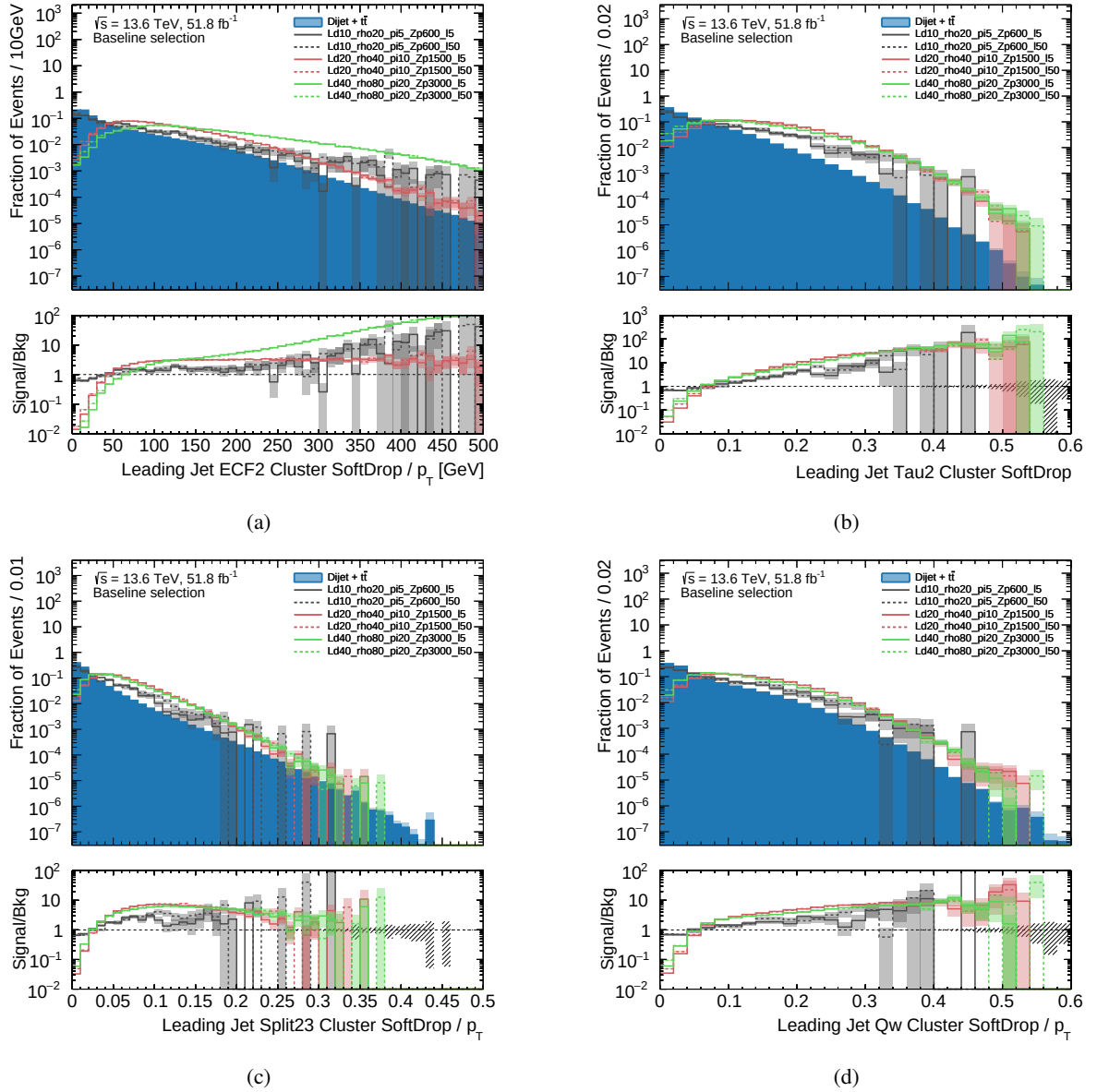


Figure 7.24: Normalised distributions of the cluster-based leading jet $ECF2/p_T$ (a), τ_2 (b), $Split23/p_T$ (c) and Q_W/p_T (d) in signal and background after the baseline selection defined in Tab. 7.4.

All observables are found to similarly discriminate signal and background. Correlations are also observed, selecting on one substructure observable removes the discriminating powers of the other observables. Hence, only one observable can be considered for the definition of the SR and the four are good candidates.

However, these observables were found to be poorly modelled in MC in the regions where a potential selection would be applied (high values). Fig. 7.25 shows the data/MC agreement for these observables. These disagreements are not well understood and thorough investigation on the potential pileup sensibility of these observables did not allow to draw any clear conclusion on their origins and on possible solutions.

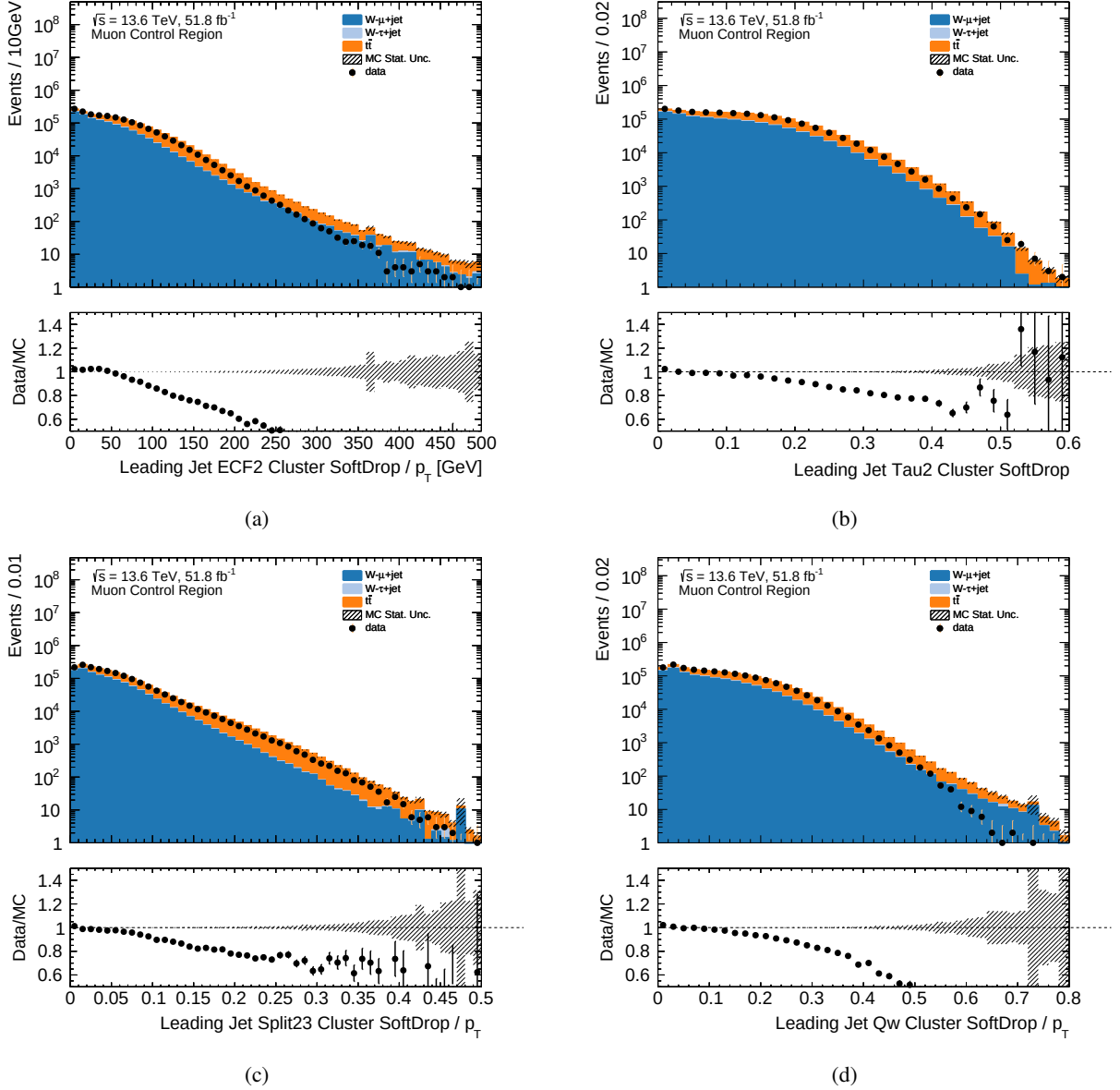


Figure 7.25: Cluster-based leading jet $ECF2/p_T$ (a), τ_2 (b), $Split23/p_T$ (c) and Qw/p_T (d) in data and MC $W+jet$ and top samples in the muon CR.

Consequently, to avoid large uncertainties coming from these discrepancies it was favoured to not use the topo-clusters as input to the computations of the substructure observables and instead use tracks within

the jets. Similarly to the topo-clusters, the SoftDrop technique is applied on the tracks to mitigate the pileup contamination. For the dimensionful observables, the jet p_T normalisation is replaced by the p_T of the corresponding track jet⁶. Fig. 7.26 shows the normalised distributions of the track-based leading jet substructure variables in signal and background and Fig. 7.27 presents the corresponding data/MC agreement plots.

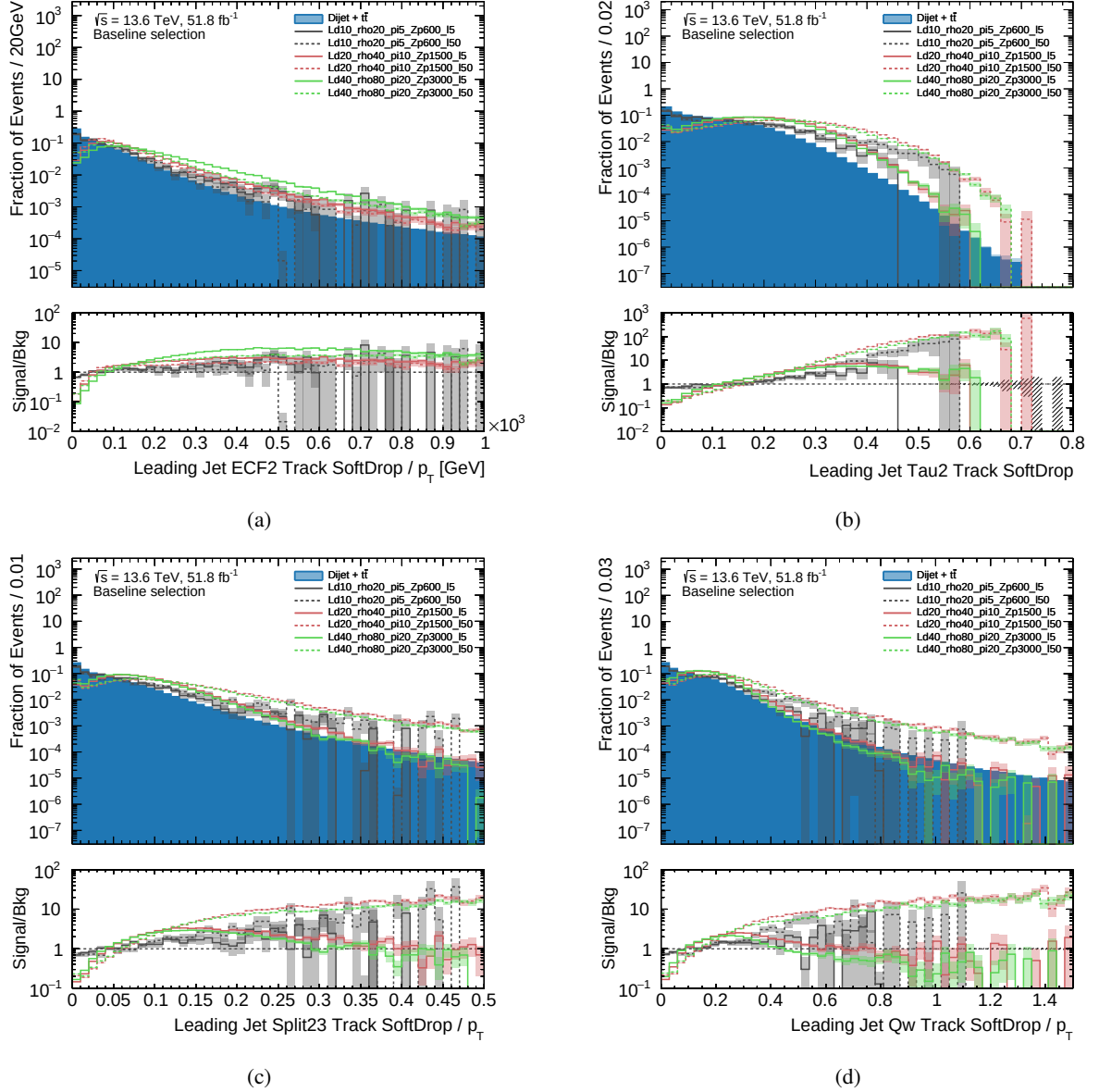


Figure 7.26: Normalised distributions of the track-based leading jet $ECF2/p_T$ (a), τ_2 (b), $Split23/p_T$ (c) and Qw/p_T (d) in signal and background after the baseline selection defined in Tab. 7.4.

The substructure observables computed from the tracks are found to be much more well-modelled in MC and in particular $ECF2$. In terms of discrimination power, similar observations are made. Moreover, the use of tracks was not found to be detrimental to the analysis as no correlations with the other track-based

⁶ The track jet p_T correspond to the p_T sum of the tracks found within the small- R jet constituents of the RC large- R jet.

observables were observed in the build up of the ABCD method, as it will be discussed in the next section.

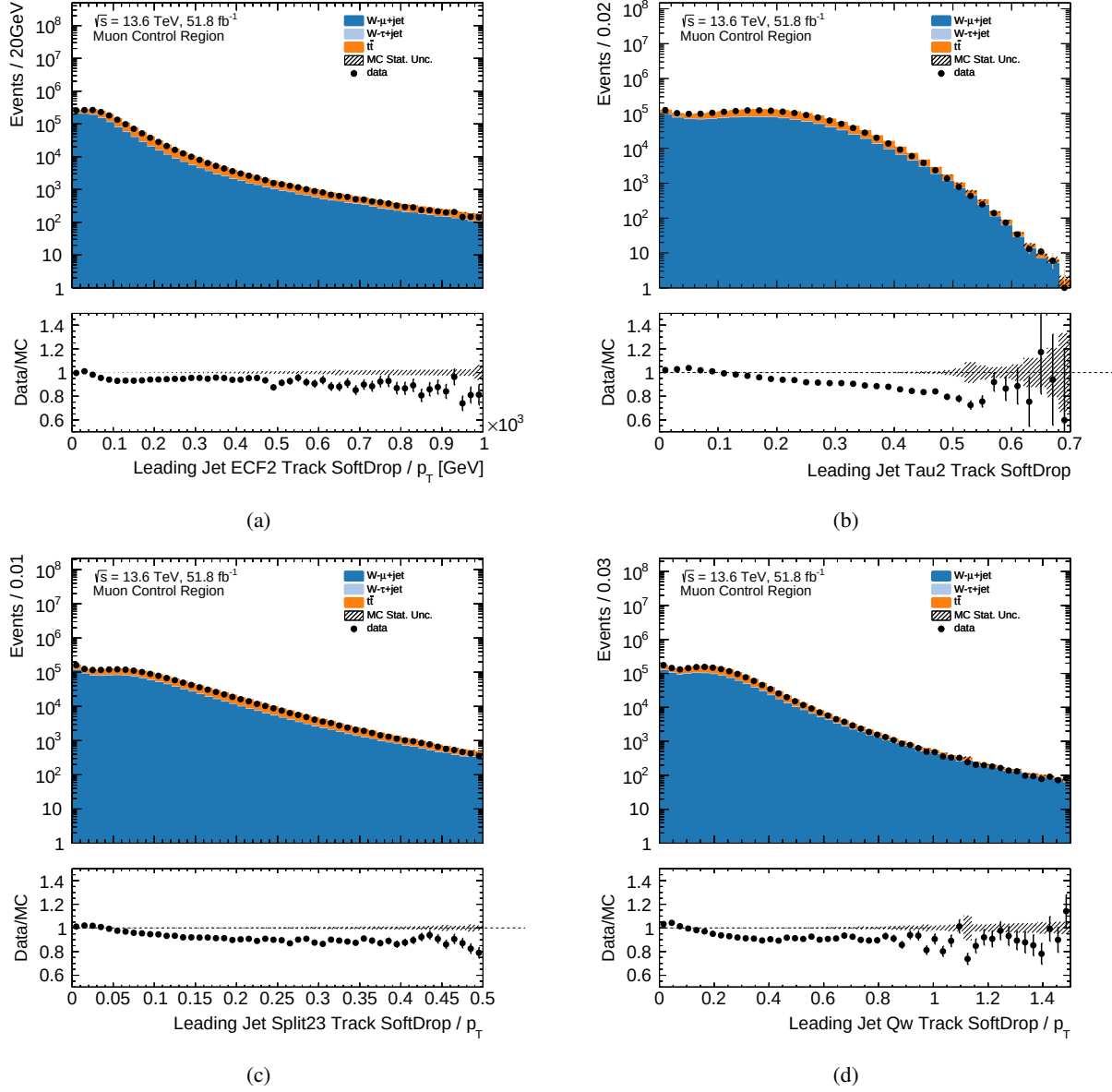


Figure 7.27: Track-based leading jet $ECF2/p_T$ (a), τ_2 (b), $Split23/p_T$ (c) and Q_w/p_T (d) in data and MC $W+\text{jet}$ and top samples in the muon CR.

Following these studies, it was decided to consider the selections on the track-based leading and/or sub-leading jet $ECF2/p_T > 40$ GeV in the definition of the SR.

7.6 Event selections: ABCD plane and signal region definitions

From the studies of the discriminating observables a pool of discriminating selections (in addition to the list of pre-selections reminded in Tab. 7.5) can be dressed and is presented in Tab. 7.6.

Table 7.5: Analysis pre-selections

Large- R jet trigger	HLT_j460_a10r_L1J100
Number of jets	≥ 2
Leading Jet	$p_T > 520$ GeV, $ \eta < 1.5$, pass OR with photon
Sub-leading Jet	$p_T > 300$ GeV, $ \eta < 1.5$, pass OR with photon
2-jet invariant mass m_{jj}	> 1 TeV

Table 7.6: Pool of discriminating observables and corresponding selections

Discriminant	Selection
Leading and/or sub-leading jet $N_{\text{vtx,sel}}$	≥ 1
Leading and/or sub-leading jet PTF	< 0.2
Leading and sub-leading jet ΔPTF	< 0.4
Leading and/or sub-leading jet track-based $ECF2/p_T$	> 40 GeV

This list of selections is used to define the ABCD plane and the SR. For the ABCD background estimation method, it is necessary to ensure that the two last selections defining the SR are made on two observables that are mostly uncorrelated for the background events. In order to find candidate pairs of uncorrelated observables, several pairs taken from the pool of discriminating observables are tested by constructing two-dimensional histograms to assess the non-correlation. The pair of observables found to be the most encouraging is the leading and sub-leading jet PTF . The 2D plane formed by the leading and sub-leading jet PTF for the background distribution is shown in Fig. 7.28 after the pre-selections presented in the previous Tab. 7.5. The background distribution is akin to a two-dimensional gaussian distribution and thus does not show a strong correlation between these two observables. While the bulk of the background distribution seems well uncorrelated, a small secondary population is observed in the bottom left corner of the plane. This secondary population can be clearly identified and observed by applying the selections $\Delta PTF \geq 0.4$ on both the leading and sub-leading jets, as shown in Fig. 7.29. These rare events were discussed in the previous section as being events with a wrongly selected HS PV. Adding the selection on the leading and sub-leading jet $\Delta PTF < 0.4$, was thus found to be necessary to remove this small subpopulation which shows a clear correlation between the leading and sub-leading jet PTF .

This 2D plane was chosen as the ABCD plane for the background estimation and the four regions (A,B,C,D) are defined using the discriminating selections on PTF :

- region A (signal enriched region) defined as leading jet $PTF < 0.2$ and sub-leading jet $PTF < 0.2$,
- region B defined as leading jet $PTF < 0.2$ and sub-leading jet $PTF > 0.2$,
- region C defined as leading jet $PTF > 0.2$ and sub-leading jet $PTF < 0.2$,
- region D defined as leading jet $PTF > 0.2$ and sub-leading jet $PTF > 0.2$.

By definition PTF is expected to fluctuate between 0 and 1 but additional tracks from pileup or the underlying event can extend the PTF distribution above 1. To avoid such effect, the ABCD plane is strictly

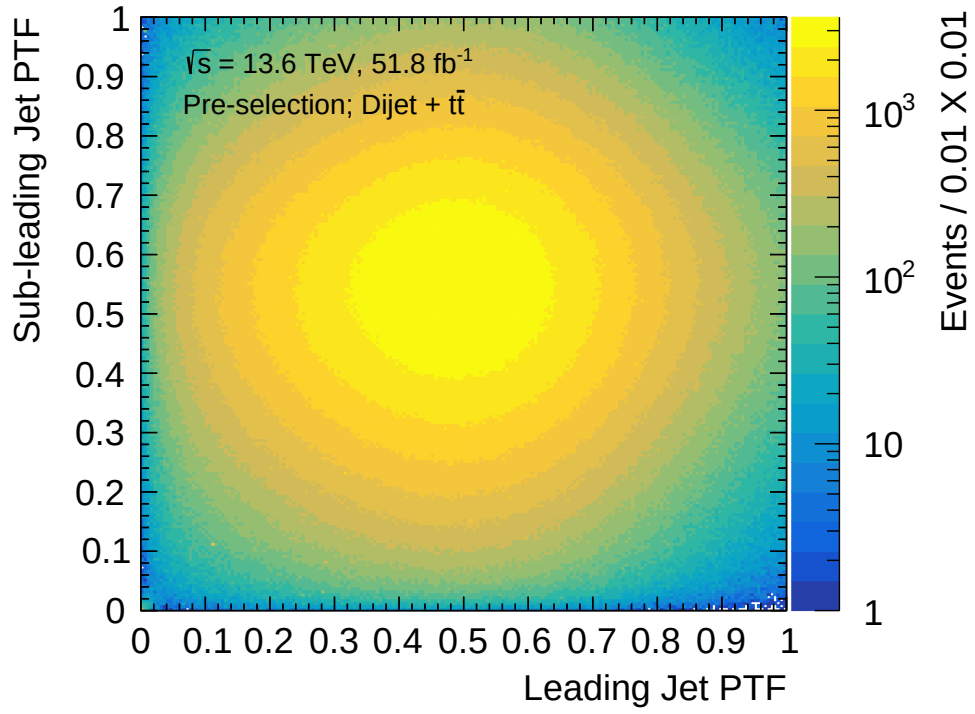


Figure 7.28: Candidate ABCD plane in background events after the pre-selections defined in Tab. 7.5.

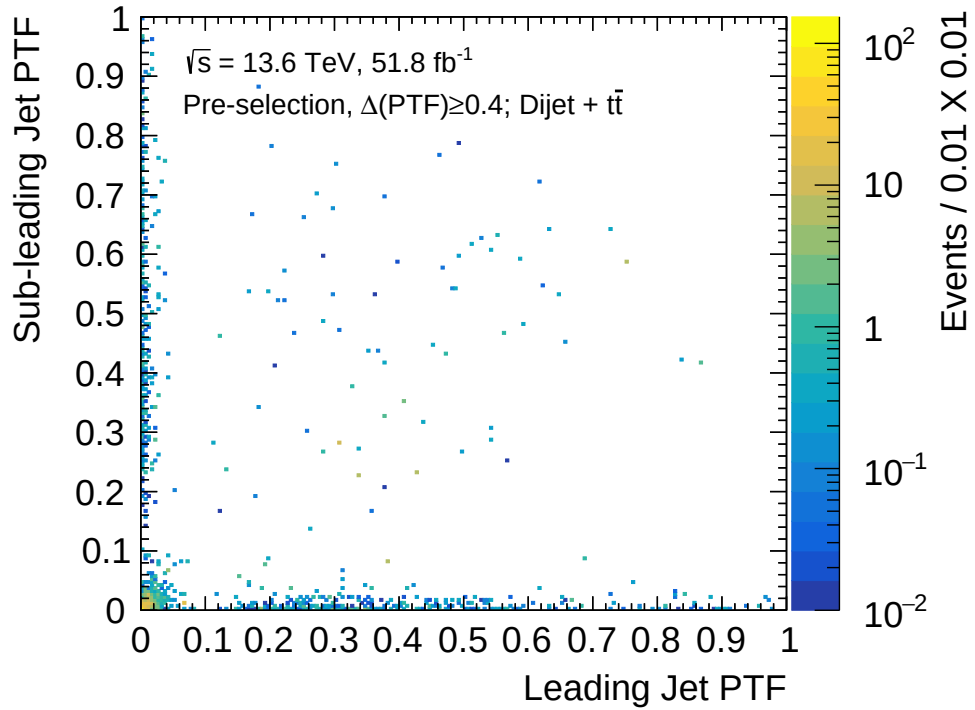


Figure 7.29: Candidate ABCD plane in background events after the pre-selections defined in Tab. 7.5 with additional selections on leading and sub-leading jet $\Delta PTF \geq 0.4$.

restricted to $0.0 \leq PTF < 1.0$ for the leading and sub-leading jets. Additionally, discriminating selections are added from the remaining pool of discriminating observables to reduce the background contribution in the region A of the ABCD plane. The powerful selection on $N_{\text{vtx,sel}} \geq 1$ is added for both the leading and sub-leading jet and only the selection on the leading jet track-based $ECF2/p_T > 40$ GeV is added. As the selection on PTF can tend to retain background events with low number of tracks, the last selection on $ECF2/p_T$ ensures that at least one of the jet in the event has a complex internal structure. The final list of selections defining the ABCD plane and the SR is given in Tab. 7.7 and the resulting ABCD plane for the background and one signal sample is shown in Fig. 7.30.

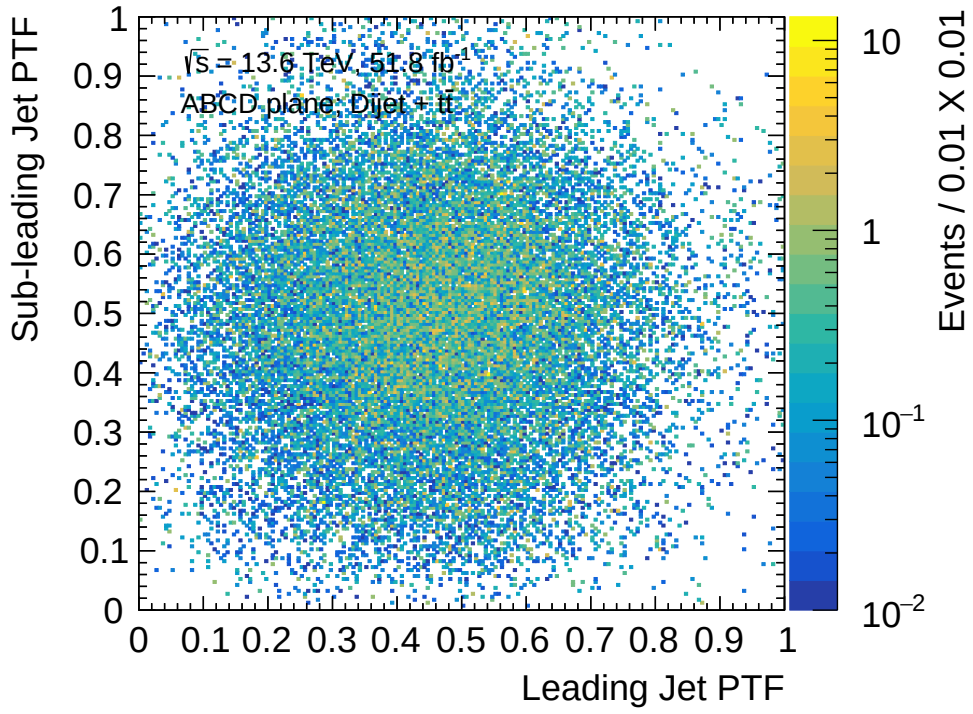
Table 7.7: ABCD plane and SR definition

Pre-selections	
Large- R jet trigger	HLT_j460_a10r_L1J100
Number of jets	≥ 2
Leading Jet	$p_T > 520$ GeV, $ \eta < 1.5$, pass OR with photon
Sub-leading Jet	$p_T > 300$ GeV, $ \eta < 1.5$, pass OR with photon
2-jet invariant mass m_{jj}	> 1 TeV
ABCD plane	
Leading and sub-leading jet ΔPTF	< 0.4
Leading and sub-leading jet $N_{\text{vtx,sel}}$	≥ 1
Leading jet track-based $ECF2/p_T$	> 40 GeV
Leading and sub-leading jet PTF	< 1.0
Signal region	
Leading and sub-leading jet PTF	< 0.2

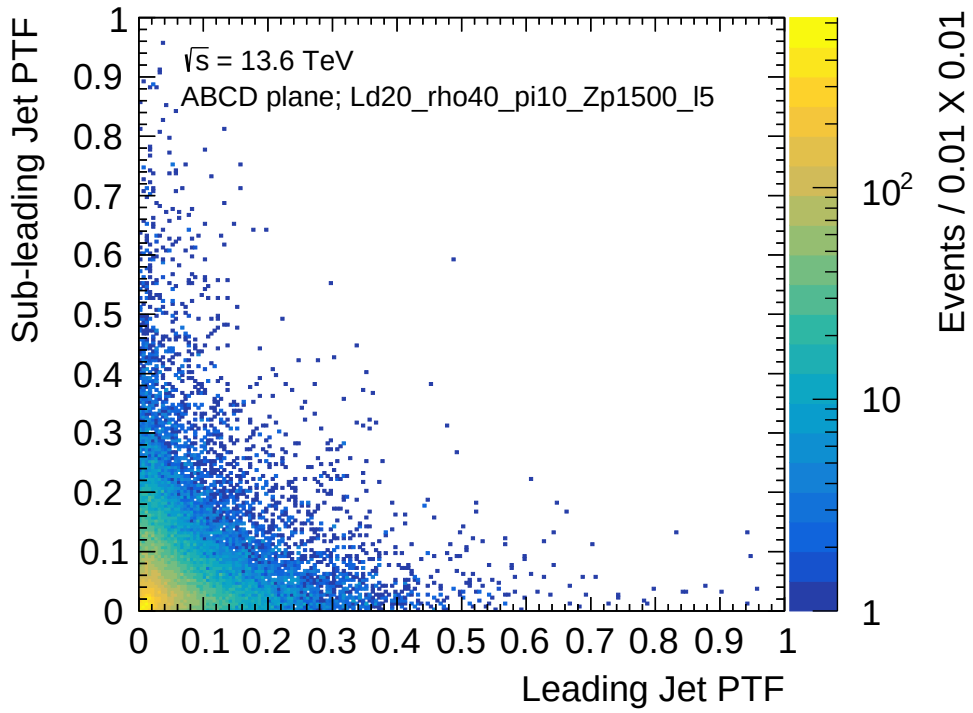
The selection cut-flow of the ABCD plane and SR is presented in Tab. 7.8 for the background and two signal samples.

Table 7.8: Analysis cut-flow. The acceptances $\mathcal{A}$ are computed with respect to the previous selection and expressed in percent.

Selection	Background (MC)		Ld20_rho40_pi10_Zp1500_15		*_150	
	N_{evt}	$\mathcal{A}$	N_{evt}	$\mathcal{A}$	N_{evt}	$\mathcal{A}$
Initial			200000		200000	
Trigger + ≥ 2 jets	112578530		135536	67.7	134892	67.5
$p_T^{\text{lead}} > 520$ GeV	55304824	49.1	103571	76.4	103448	76.7
$p_T^{\text{sub-lead}} > 300$ GeV	52825973	95.5	102988	99.4	100098	96.8
$ \eta < 1.5$	33424814	63.3	86574	84.1	84305	84.2
Overlap removal	33290808	99.6	86379	99.8	83992	99.6
$m_{jj} > 1$ TeV	30839861	92.6	84906	98.3	81509	97.0
$\Delta PTF < 0.4$	30791400	99.8	84513	99.5	81151	99.6
$N_{\text{vtx,sel}}^{\text{lead}} \geq 1$	488704	1.59	70043	82.9	50884	62.7
$N_{\text{vtx,sel}}^{\text{sub-lead}} \geq 1$	8794	1.80	58798	84.0	33302	65.5
$ECF2/p_T > 40$ GeV	5879	66.9	54194	92.2	29433	88.4
$PTF < 1.0$	5763	98.0	54151	99.9	29418	99.9
SR ($PTF < 0.2$)	27	0.46	48307	89.2	28131	95.6



(a)



(b)

Figure 7.30: Final ABCD plane for background (a) (the limit on the z-axis has been lowered to 10^{-2}) and the signal sample Ld20_rho40_pi10_Zp1500_l5 (b). All analysis selections of Tab. 7.7 up to the signal region are applied.

Very few background events are expected to be retained in the ABCD plane (≈ 6000) and only 27 events are predicted in the SR. A more in detail table of the number of events in each region of the ABCD plane and for each source of background is given in Tab. 7.9. The signal, on the other hand, is well-preserved and most of the events are found in the SR (89% of the events of the ABCD plane for the model Ld20_rho40_pi10_Zp1500_15).

Table 7.9: Number of MC events in each region of the ABCD plane for the different sources of background (dijet and $t\bar{t}$).

	Dijet	$t\bar{t}$	Full background
A	26	1	27
B	402	8	410
C	345	7	352
D	4879	95	4974
Total	5652	111	5763

The total signal acceptance in the SR $\mathcal{A}_{\text{tot}}$ and the signal leakages q_X ($N_{\text{sig}}^X/N_{\text{sig}}^A$) in the CRs of the ABCD plane are given in Tab. 7.10 for all models. The SR shows an excellent signal acceptance ranging from 7% and up to around 50% for the models of interest for this analysis scenario. The signal leakages q_X in the CRs are also very good with only a few percent signal losses. For the same dark QCD parameters an increase in acceptance with the mass of the Z' is observed and is mainly the result of the higher acceptance of the selections on the leading and sub-leading jet p_T . The loss of acceptance at higher lifetimes, when going from models with $c\tau_{\pi_D} = 5$ mm to 50 mm, is primarily caused by the decrease in the vertex efficiency reconstruction. Similarly, the gain in acceptance when increasing the mass of the dark pion is the outcome of the increase in vertex efficiency reconstruction, as observed in Fig. 7.7.

Table 7.10: Signal acceptance and leakages $q_X = N_{\text{sig}}^X/N_{\text{sig}}^A$ in percent.

Signal model	$\mathcal{A}_{\text{tot}}$	q_B	q_C	q_D
Ld10_rho20_pi5_Zp600_15	0.002	1132	738	84.1
Ld10_rho20_pi5_Zp600_150	0.001	206	559	72.1
Ld10_rho20_pi5_Zp1500_15	18.0	6.24	4.42	0.25
Ld10_rho20_pi5_Zp1500_150	7.3	2.35	2.33	0.02
Ld10_rho20_pi5_Zp3000_15	39.2	3.27	2.37	0.08
Ld10_rho20_pi5_Zp3000_150	19.2	1.42	1.00	0.02
Ld20_rho40_pi10_Zp600_15	0.006	551	444	90.1
Ld20_rho40_pi10_Zp600_150	0.005	412	288	34.2
Ld20_rho40_pi10_Zp1500_15	24.2	6.78	4.97	0.35
Ld20_rho40_pi10_Zp1500_150	14.1	2.67	1.87	0.03
Ld20_rho40_pi10_Zp3000_15	49.0	4.06	3.19	0.12
Ld20_rho40_pi10_Zp3000_150	31.5	1.43	0.99	0.02
Ld40_rho80_pi20_Zp600_15	0.011	325	213	34.9
Ld40_rho80_pi20_Zp600_150	0.003	750	526	72.1
Ld40_rho80_pi20_Zp1500_15	23.9	8.08	5.78	0.45
Ld40_rho80_pi20_Zp1500_150	15.5	2.95	1.81	0.07
Ld40_rho80_pi20_Zp3000_15	48.0	5.24	4.08	0.21
Ld40_rho80_pi20_Zp3000_150	32.3	1.52	1.01	0.01

7.7 Background estimation, validation of the ABCD method

The ABCD plane and SR were defined in the previous section and are employed in this section to validate the ABCD background estimation method. This section shows, in MC, that the ABCD method is able to estimate the right background yield in the SR. The validation in MC is then extended to VRs outside the SR to ensure that the background distribution is sufficiently uncorrelated in the full ABCD plane and that the ABCD method consistently provides good background estimations. Finally, the background estimation method is similarly validated in data, in subregions of the ABCD plane that are expected to be dominated by background events and mostly free from signal contamination. All the validations are made using the standard approach of the ABCD method, as discussed in Sec. 7.4.

7.7.1 Nominal test of the ABCD background estimation method in MC

The first validation of the ABCD method is made in MC on the nominal ABCD plane and defined regions. Fig. 7.31 presents the ABCD plane for the MC background events with the defining boundaries of the four ABCD regions.

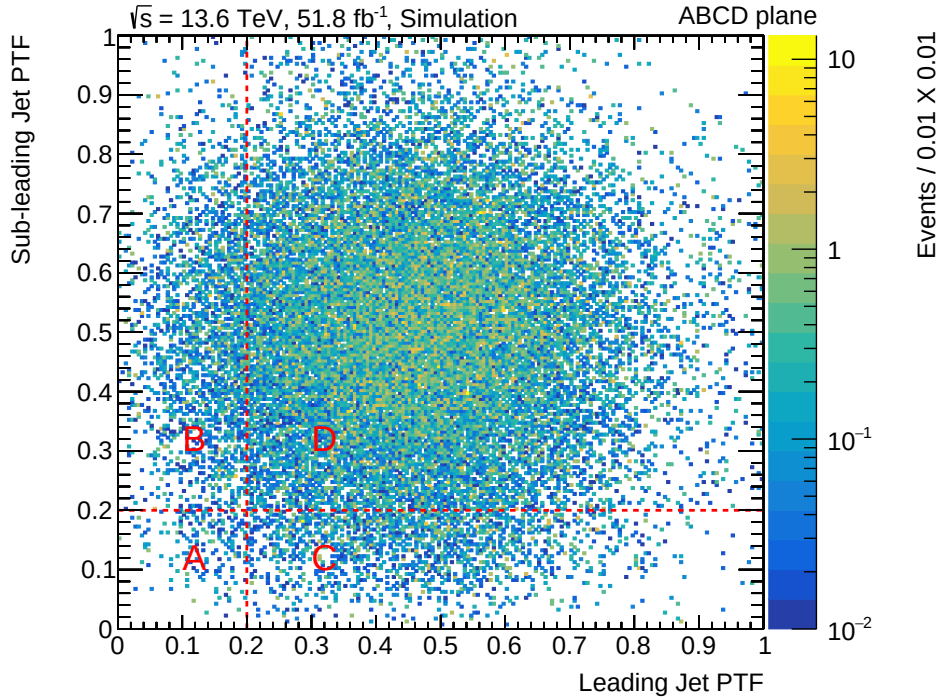


Figure 7.31: ABCD plane with the nominal SR and CRs containing MC background events.

The observed number of background events and statistical uncertainties in each region are given in Eq. 7.15. The MC background samples being re-weighted, as discussed in Sec. 7.1, the numbers of events per region are thus not integers.

$$\begin{aligned}
N_{\text{bkg}}^A &= 27.2 \pm 5.2 \\
N_{\text{bkg}}^B &= 409.8 \pm 20.2 \\
N_{\text{bkg}}^C &= 352.3 \pm 18.8 \\
N_{\text{bkg}}^D &= 4974.4 \pm 70.5
\end{aligned} \tag{7.15}$$

The ABCD estimation of the number of background events in region A and its statistical uncertainty are then simply obtained by applying Eq. 7.9 and are given in Eq. 7.16.

$$\begin{aligned}
N_{\text{bkg,estimated}}^A &= \frac{N_{\text{bkg}}^C}{N_{\text{bkg}}^D} N_{\text{bkg}}^B \\
N_{\text{bkg,estimated}}^A &= 29.0 \pm 2.1
\end{aligned} \tag{7.16}$$

The estimated number of background events in the region A (SR) is thus shown to be consistent with the observed value as the uncertainties largely cover the difference between the observation and estimation.

7.7.2 Validation of the ABCD background estimation method in MC

An extensive validation procedure is applied on MC to ensure that the region definitions are not a particular case for which the ABCD relationship between the number of background events in each region is valid. As discussed in Sec. 7.4, this validation consists in considering subregions of the ABCD plane excluding the SR and in which several VRs and corresponding new CRs can be defined. The ABCD method is then applied to each VR to validate that the method consistently provides good background estimations in various areas of the ABCD plane. An example of such SR exclusion and VR and CRs definitions is given in Fig. 7.32 where the new regions are indicated with a "" and the excluded region is darkened. Two validation configurations are considered: one where the SR is excluded by restricting the ABCD plane with a lower cut on the leading jet *PTF* at 0.2, and one restricting the plane with a lower cut on the sub-leading jet *PTF* at 0.2. In both cases, the VR is defined as the region the closest to the nominal SR (bottom left corner) and one of the defining selection of the SR (the one not used to exclude the SR) is retained. These two last considerations allow to test the ABCD method in configurations close to the nominal ABCD plane. Finally, several VRs and corresponding CRs are considered by varying the unrestricted selection on the leading or sub-leading jet *PTF*.

The first configuration is illustrated in Fig. 7.32, the SR is excluded with a lower cut on the leading jet *PTF*, the selection on the sub-leading jet *PTF* is retained and the final free VR defining selection on the leading jet *PTF* is, in this example, set to 0.4. Variations on the leading jet *PTF* selection are considered between 0.2 and 0.8 with steps of 0.01. For each varied selection, the expected number of background events in the VR is estimated with the ABCD method and compared to the observed value. The results are reported in Fig. 7.33 showing as a function of the leading jet *PTF* cut, the observed and expected number of events in the VR or region A'. The ABCD expectations are presented as an error band. For all the set of tested VRs, an excellent agreement between the observations and ABCD estimations is observed.

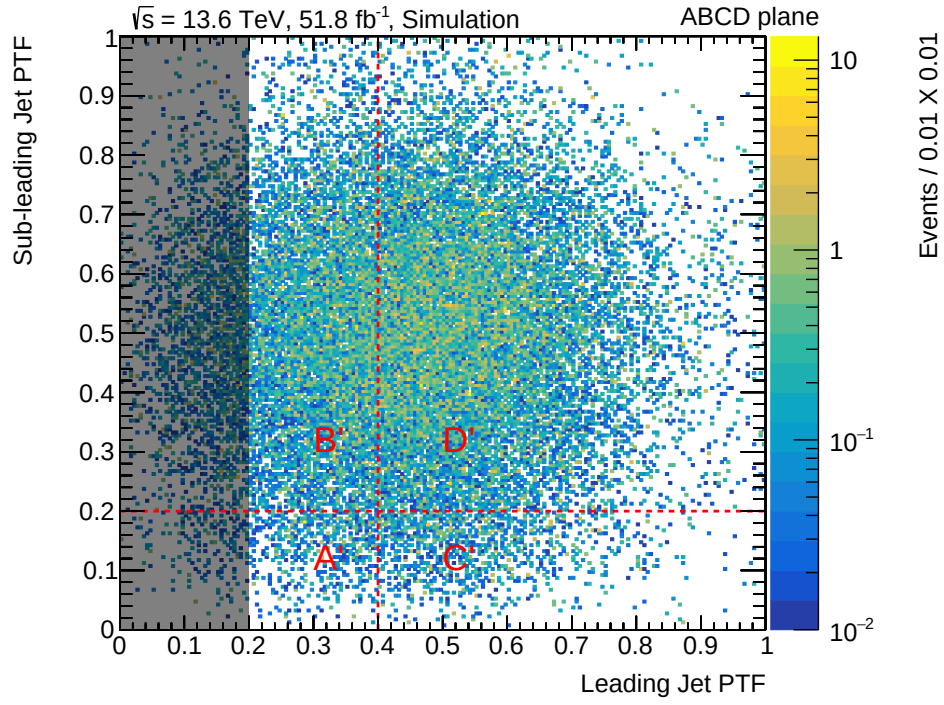


Figure 7.32: ABCD plane containing MC background events and validation regions A', B', C' and D'. The darkened area represents the excluded region.

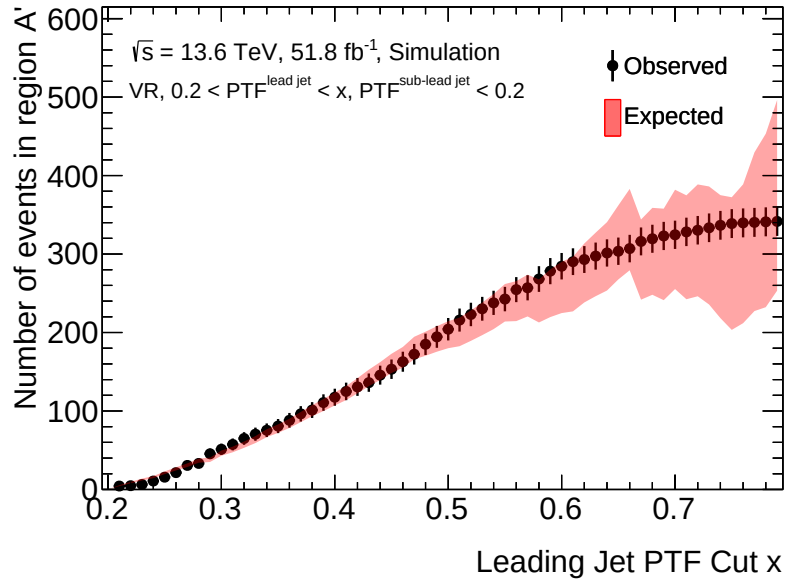


Figure 7.33: Observed and expected number of events in the validation regions A' as a function of the leading jet *PTF* selection cut.

The second configuration is similarly illustrated in Fig. 7.34 and is the complete mirror of the first configuration, where the leading and sub-leading jet PTF roles are inverted. The results of the validation of the ABCD method for this configuration are shown in Fig. 7.35 showing the expected and observed number of events in each VR as a function of the sub-leading jet PTF cut. As for the first validation, the agreement between observations and expectations is very good for a large set of VRs.

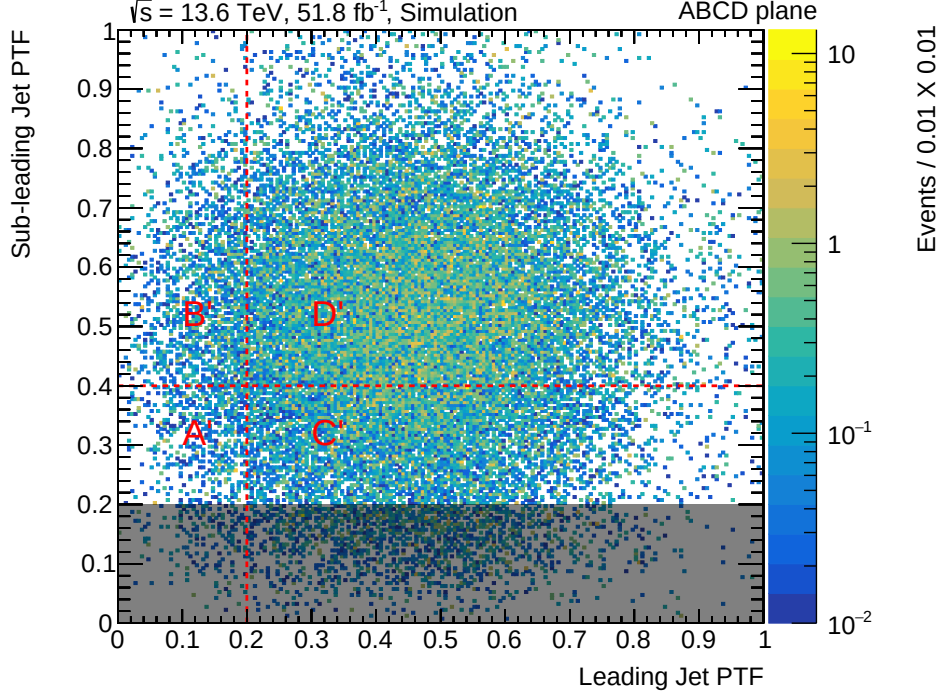


Figure 7.34: ABCD plane containing MC background events and validation regions A', B', C' and D'. The darkened area represents the excluded region.

A very small non-closure is observed for selections on the sub-leading jet PTF around 0.7. This is caused by the migration of localised groups of events from a region to another when varying the definitions of the VR and CRs. These groups, when fully contained in one region, do not bias the ABCD estimations but, when migrating to other regions, can produce such small non-closure. It can be a sign of over-densities that could reveal some correlation between the leading and sub-leading jet PTF in very specific subregions of the ABCD plane. Alternatively, it possibly is just an artefact of the MC simulation and the re-weighting scheme of the dijet samples that generate one event with a very large MC weight. The generation of the dijet samples in slices of p_T in rare cases can create events with a leading jet (usually from pileup) that wander off the expected p_T range of a slice and end up in higher- p_T slices. Such events are, however, still generated with the MC weights corresponding to the generated slice. As the MC weight are used to scale the number of generated events to the cross-section and luminosity, lower p_T -slices events have higher MC weights than the larger p_T -slices. An event from a low- p_T slice with a larger than expected leading jet p_T will end up in the p_T range of higher slice with a MC weight much higher than it should. These kinds of events then can pass analysis selections and create large peaks in observable distributions⁷.

⁷ This phenomenon was especially observable in Fig. 7.20 and 7.23. These figures also shown a correlation between events with a wrongly selected PV and events with a large MC weights. Both are the consequence of a large- p_T pileup jet biasing the PV

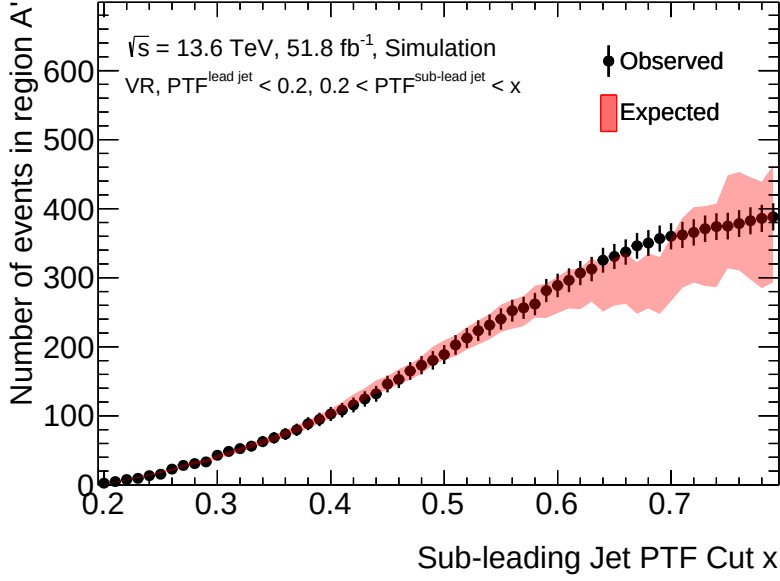


Figure 7.35: Observed and expected number of events in the validation regions A' as a function of the sub-leading jet PTF selection cut.

Selections are applied to remove such events but they are, however, not always fully efficient. Anyway, this small non-closure is not seen as a strong showstopper as it is a very limited discrepancy in an ABCD configuration far off from the nominal ABCD plane and SR. Furthermore, the ABCD method needs to be primarily validated in data where this kind of issue is not present.

7.7.3 Validation of the ABCD background estimation method in blinded data

The validation of the ABCD method in MC provides a good sense of the consistency of the method to estimate the background contribution in the SR. The final validity of the method, however, needs to be assessed in data. In data, contrary to MCs, it is obviously not possible to separate background from a potential signal as it is the purpose of the analysis. Thus, it is not possible to validate the background estimation method directly in the nominal SR as it could contain signal. The SR is thus blinded in data. Additionally, the CRs were also shown to contain a few percent of the hypothesised signal, so in order to avoid potential signal contamination in data, a wider region than the SR is actually blinded in data. The region with leading or sub-leading jet $PTF < 0.4$ is excluded, as signal outside this region is very much negligible compared to the dominant SM background. The resulting truncated ABCD plane for data events is shown in Fig. 7.36.

selection and escaping the intended p_T range of the generated slice.

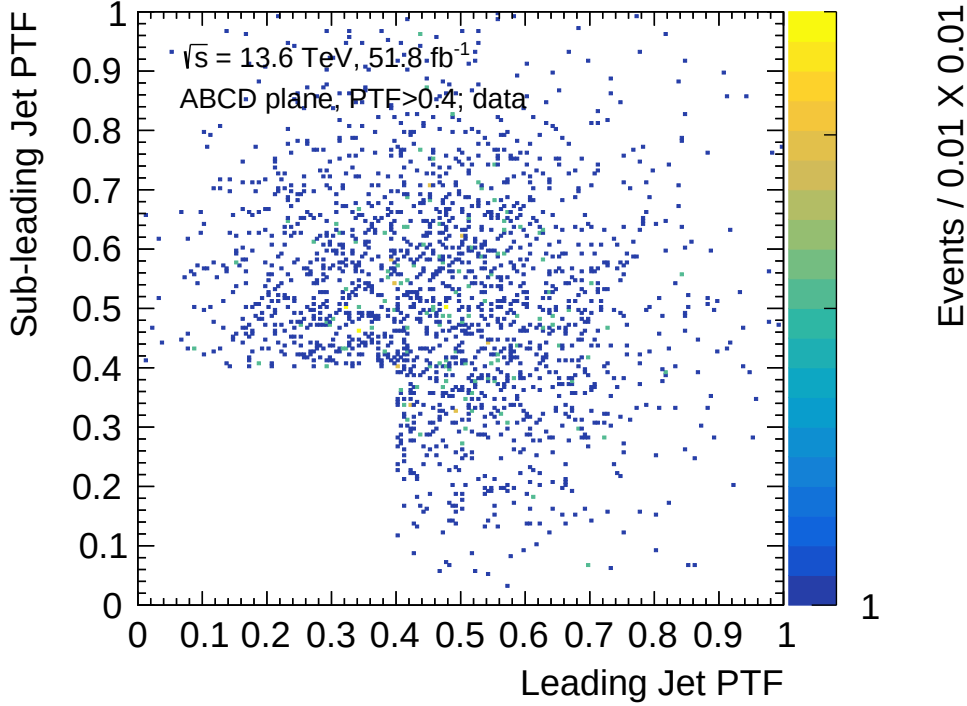
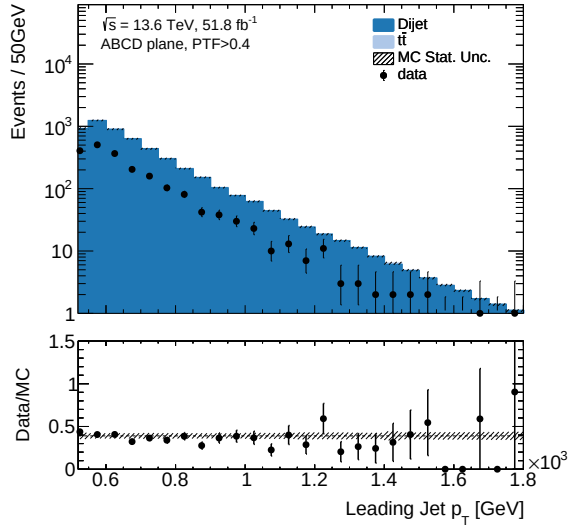


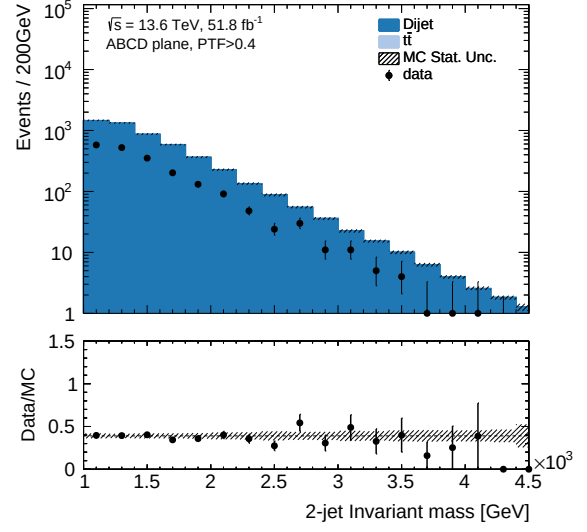
Figure 7.36: ABCD plane containing data with additional selection on leading or sub-leading $PTF > 0.4$.

7.7.3.1 Data/MC agreement in the ABCD validation region

As a sanity check, an additional study of the agreement between data and the MC background in the ABCD blinded validation region is made to ensure the good modelling of the observables of interests and of the analysis selections. Better agreement in the ABCD validation region between data and the actual MC background are expected compared to the previous data/MC agreement studies presented in Section 7.5 as the $t\bar{t}$ contribution is largely negligible in this ABCD validation region. Fig. 7.37, 7.38 and 7.39 present the data/MC agreement in such a region and with the MC background samples for the main observables used to define the analysis selections. Other than a difference in normalisation, the data/MC ratios are shown to be very good with almost flat curves for most of the observables. Improvements are especially observed for the leading jet p_T and m_{jj} distributions compared to Fig. 7.14 and 7.15, validating the fact that the previously observed mis-modelling is linked to the $t\bar{t}$ contribution in the muon CR, a process which is a negligible source of background in this analysis. Mis-modelling are still observed for ΔPTF and $N_{vtx,sel}$. The blinded analysis cutflow in data and MC background given in Tab. 7.11 shows that the selection acceptances are consistent between data and MC for all selections except the ones on $N_{vtx,sel}$. This shows, in particular, that the mis-modelling of the distribution of ΔPTF , firstly observed in Sec. 7.5 and supposed to be not problematic, is confirmed to have no effect on the data/MC acceptance ratio (the distribution integral between 0 and 0.4 is similar between data and MC). On the other hand, as discussed before, the mis-modelling of $N_{vtx,sel}$ is expected from the material map veto and not considered as a difficulty for the following statistical analysis. Indeed, the signal acceptance should not be affected as the signal vertices are not preferentially produced in the tracker material as it is the case for the background.

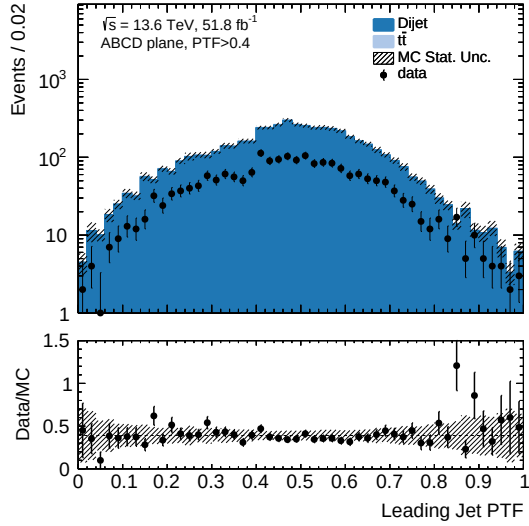


(a)

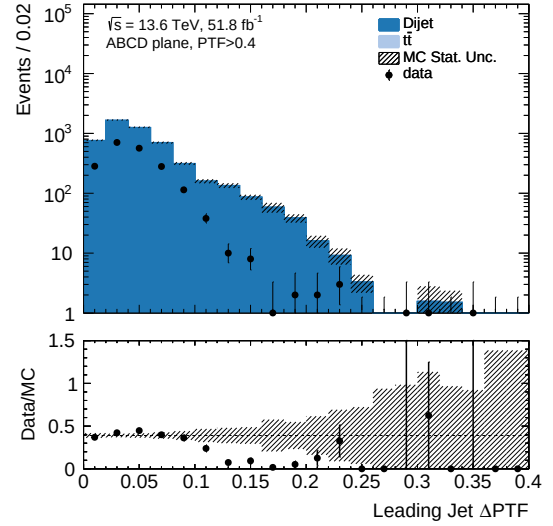


(b)

Figure 7.37: Leading jet p_T (a) and 2-jet invariant mass (b) in data and MC dijet and top samples in the ABCD blinded plane (blinding with leading or sub-leading jet $PTF > 0.4$).



(a)



(b)

Figure 7.38: Leading jet PTF (a) and leading jet ΔPTF (b) in data and MC dijet and top samples in the ABCD blinded plane (blinding with leading or sub-leading jet $PTF > 0.4$).

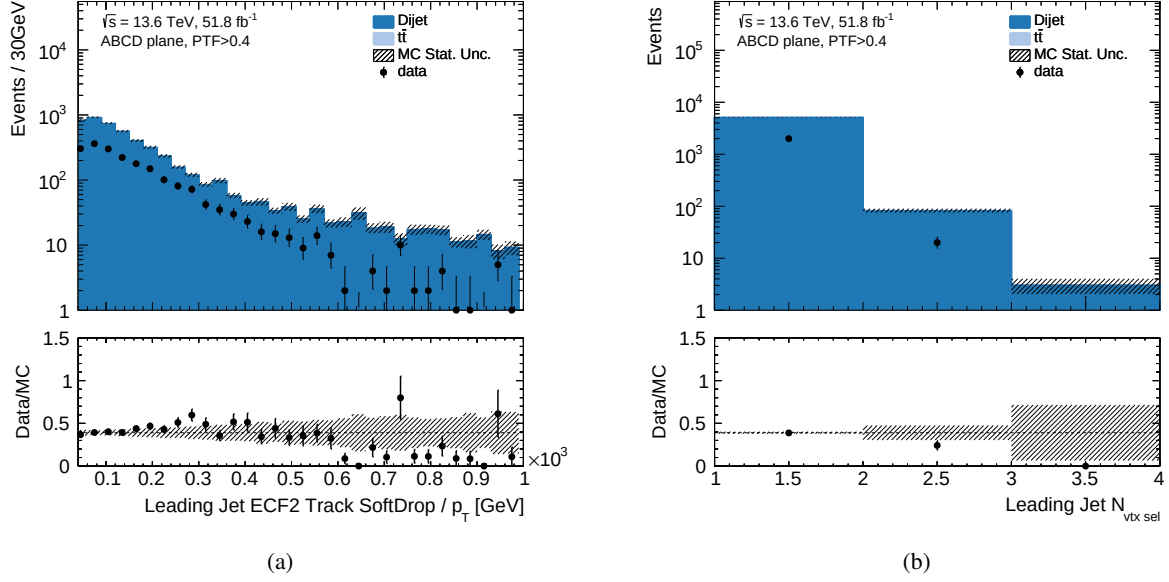


Figure 7.39: Leading jet track-based $ECF2/p_T$ (a) and leading jet $N_{vtx,sel}$ (b) in data and MC dijet and top samples in the ABCD blinded plane (blinding with leading or sub-leading jet $PTF > 0.4$).

Table 7.11: Blinded analysis cutflow in data and MC background. The acceptances $\mathcal{A}$ are computed with respect to the previous selection and expressed in percent.

Selection	Background (MC)		Data		Data/MC	
	N_{evt}	$\mathcal{A}$	N_{evt}	$\mathcal{A}$	N_{evt}	$\mathcal{A}$
Trigger + ≥ 2 jets	112578530		95119321		0.84	
Blinding ($PTF > 0.4$)	102753231	91.3	84919757	89.3	0.83	0.98
$p_T^{lead} > 520$ GeV	49949309	48.6	37749478	44.5	0.76	0.91
$p_T^{sub-lead} > 300$ GeV	47718502	95.5	36388445	96.4	0.76	1.01
$ \eta < 1.5$	30771398	64.5	23706597	65.2	0.77	1.01
Overlap removal	30658768	99.6	23574869	99.4	0.77	1.00
$m_{jj} > 1$ TeV	28378538	92.6	21858417	92.7	0.77	1.00
$\Delta PTF < 0.4$	28337014	99.9	21818100	99.8	0.77	1.00
$N_{vtx,sel}^{lead} \geq 1$	444328	1.57	234538	1.07	0.53	0.69
$N_{vtx,sel}^{sub-lead} \geq 1$	7792	1.75	2836	1.21	0.36	0.69
$ECF2/p_T > 40$ GeV	5231	67.1	2016	71.1	0.39	1.06
$PTF < 1.0$	5116	98.0	1970	97.7	0.39	1.00

7.7.3.2 Validation of the ABCD method in data

As in MC, in data the ABCD method can be validated in VRs. To obtain complete regions and avoid the excluded region, two subregions of the ABCD plane are considered: one excluding the region with the sub-leading jet $PTF < 0.4$ (illustrated in Fig. 7.40) and a second excluding the region with the leading jet $PTF < 0.4$ (illustrated in Fig. 7.42). In each subregion, the ABCD method is validated in several VRs by varying both selections defining the VRs and comparing the observed and ABCD-expected number of events. Again, the VRs are always defined as the region with lowest PTF (bottom left corner).

In the configuration illustrated by Fig. 7.40, the selection on the leading jet PTF is varied between 0.1 and 0.5 with 0.05 steps and the selection on the sub-leading jet PTF is varied between 0.5 and 0.8 with steps of 0.1. The results are presented in Fig. 7.41 as a function of the leading jet PTF cut and for each selection on the sub-leading jet PTF . Good agreement between the observed and expected values are observed for all the VRs.

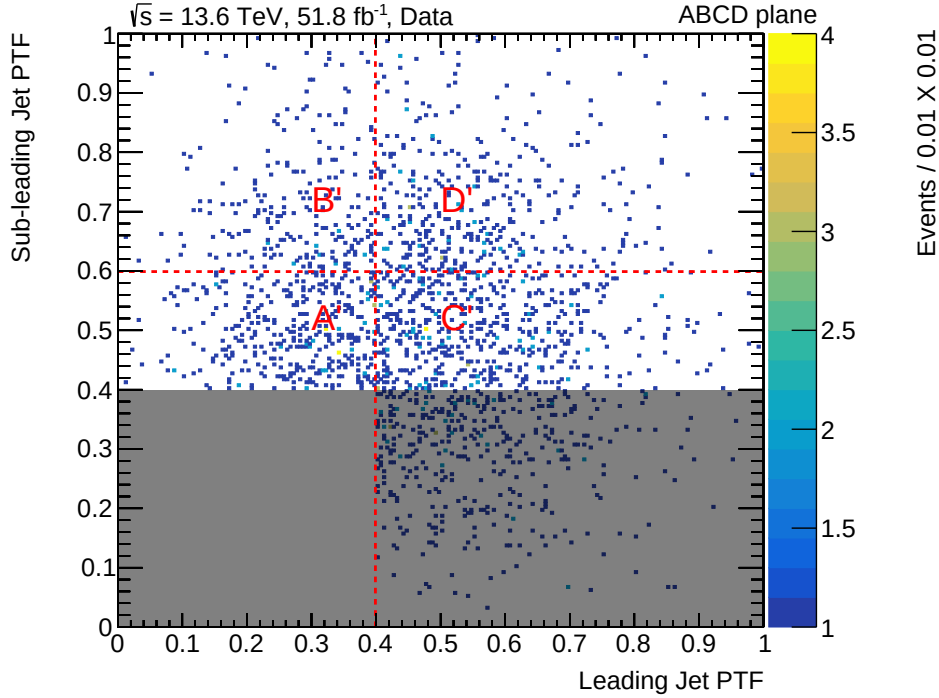


Figure 7.40: ABCD plane containing data with additional selection on leading or sub-leading $PTF > 0.4$ and validation regions A', B', C' and D'. The darkened area represents the excluded region.

In the second configuration illustrated in Fig. 7.42, the same procedure is applied by inverting the roles of the leading and sub-leading jet PTF and the results are presented in Fig. 7.43. Similar observations are made on the agreement between observations and expectations on the number of events in the VRs.

The ABCD background estimation method is validated in both MC and blinded data. It is shown to be consistently providing precise estimations on the number of background events for a large set of ABCD configurations. It is thus assumed that this method with the discussed analysis selections, ABCD plane definition and SR definition will be capable, in data, of predicting precisely the expected background

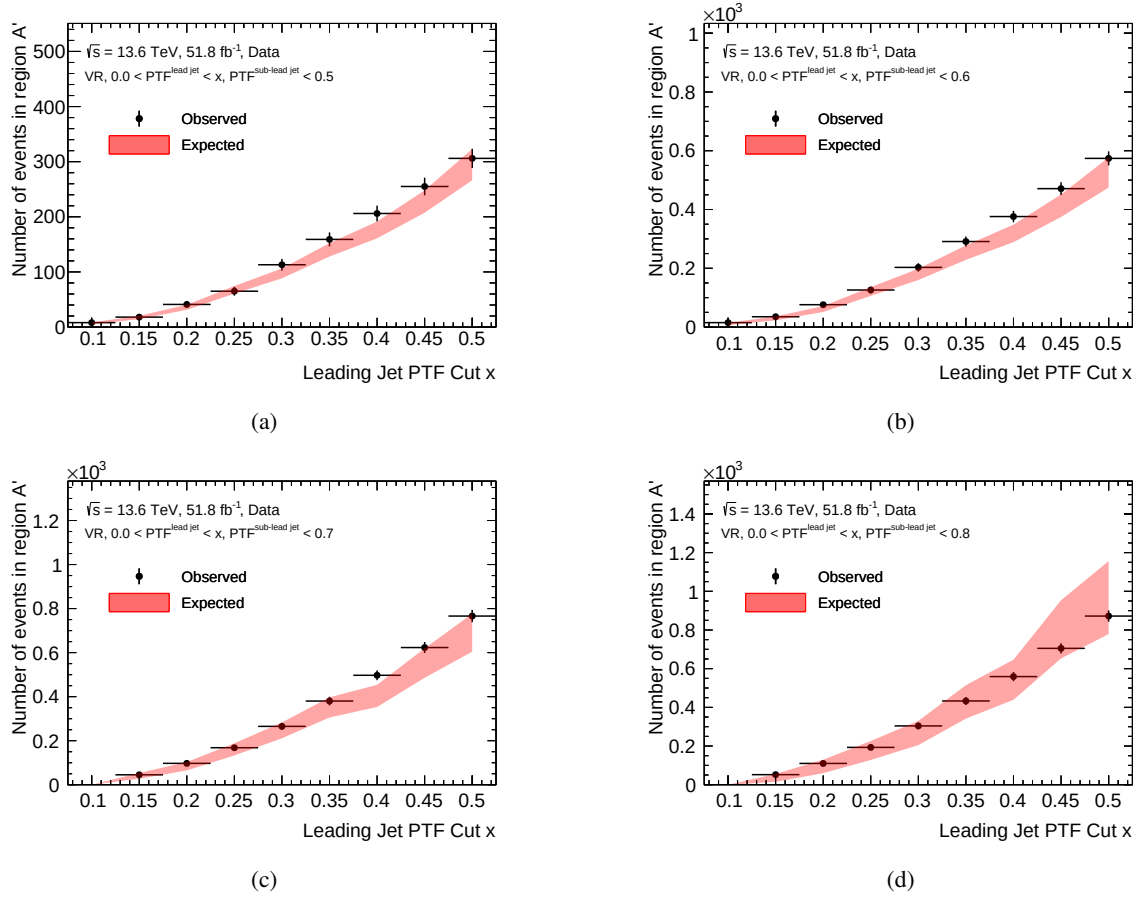


Figure 7.41: Observed and expected number of events in the validation regions A' as a function of the leading jet PTF selection cut.

contribution in the SR. The analysis strategy is therefore validated and can be used for the development of the statistical analysis.

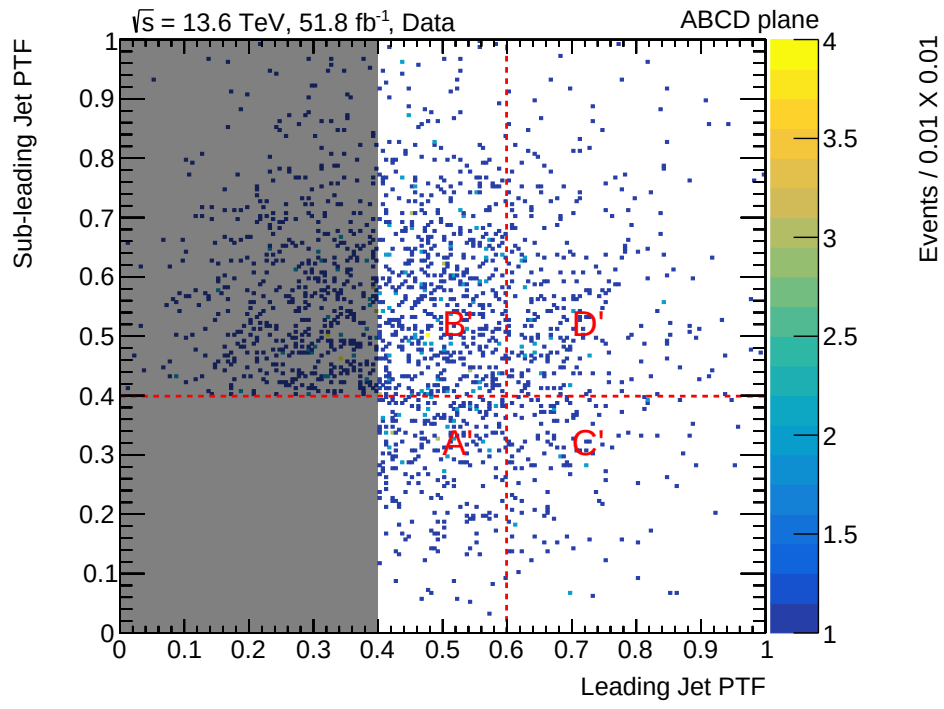
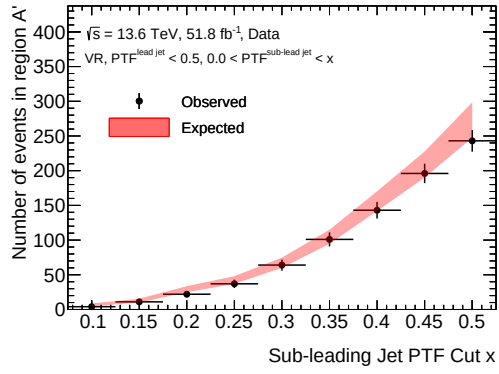
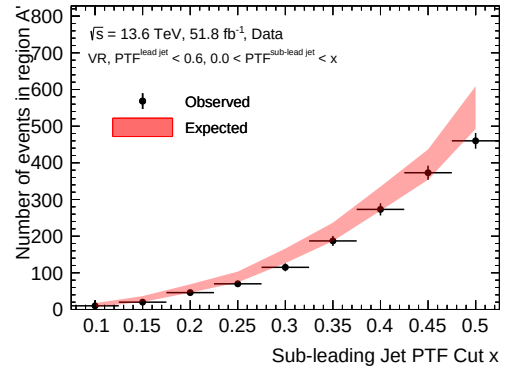


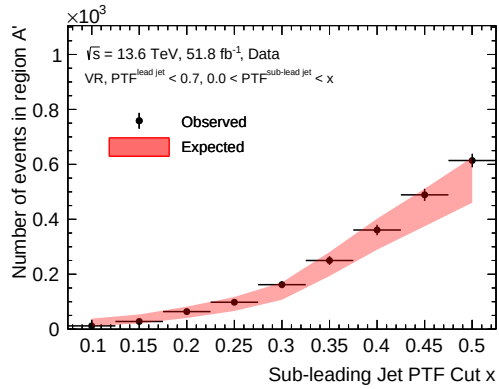
Figure 7.42: ABCD plane containing data with additional selection on leading or sub-leading $PTF > 0.4$ and validation regions A, B, C and D. The darkened area represents the excluded region.



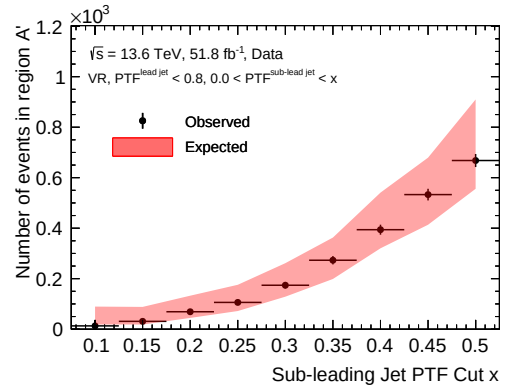
(a)



(b)



(c)



(d)

Figure 7.43: Observed and expected number of events in the validation regions A' as a function of the sub-leading jet PTF selection cut.

7.8 Systematic uncertainties

Systematic uncertainties in the MC signal yields in the ABCD plane regions and thus in the acceptances need to be considered for the statistical analysis. These uncertainties correspond to variations between simulation and the real values or distributions of the observables used to define the ABCD plane and SR which lead to variations of the signal yields. As the analysis background estimation method is data-driven and has been validated in the previous section, no systematics are applied to the background estimation.

Experimental systematic uncertainties from the reconstruction of various objects are taken into account. In the analysis selections, observables built on jets, tracks and vertices are used. Systematics on the jet, track and vertex reconstruction are considered and discussed in this section. Additionally, theoretical uncertainties in the MC generator are also considered through the variations of the MC weights (the signal samples are re-weighted as discussed in Sec. 7.1). A systematic uncertainty in the pileup re-weighting scheme is also included and discussed in the following. Each systematic variation is propagated through the analysis selections to obtain a systematic variation on the signal yield in each region of the ABCD plane.

7.8.1 Tracking systematics

The tracking systematics take into account the tracking reconstruction efficiency and the rate of fake tracks. These are used to evaluate the effect of systematics by randomly dropping out tracks from the pool of tracks. Different systematics are derived for the standard and LRT tracks. For standard tracks, the uncertainties in the reconstruction efficiency are assessed by comparing the efficiency of track reconstruction in MC samples with different detector configurations varying the tracker passive material. The rate of fake tracks provides an additional uncertainty coming from the potential combinatorial fakes of the tracker hits. This rate is derived in data, based on the observation of non-linearities in the distribution of the per-event track multiplicity as a function of $\langle\mu\rangle$. As the track multiplicity is expected to scale linearly with $\langle\mu\rangle$, it is assumed that the non-linearities are produced by fake tracks. Due to the excellent performance of the inner tracker, the tracking systematic uncertainties for standard tracks are dominated by the tracking fake rate. For LRT tracks, only systematics on the reconstruction efficiency are recommended. They are derived from comparisons between data and MC of the tracks produced by the displaced decays of kaons. Additional uncertainties are derived for tracks in dense environment (i.e. jet cores) in which the track density can overcome the granularity of the tracker and lead to clusters of energy deposits compatible with multiple particles but associated with a single track. This leads to additional efficiency and fake uncertainties that are only applied to tracks within $\Delta R = 0.1$ of a jet. Each systematic is applied on the pool of tracks by randomly dropping tracks with a defined probability corresponding to the efficiency or fake rate uncertainty. The obtained systematically varied collections of tracks are then used to compute the systematic variations of the track-based observables (PTF , ΔPTF , and $ECF2/p_T$).

The track systematics are also propagated to the vertex reconstruction to obtain systematically varied collections of vertices. Ideally, each tracking systematic and the corresponding varied track collection should be used to produce a varied collection of vertices by re-applying the vertex reconstruction algorithm on the varied track collection. However, for simplicity and gain in computing resources all the track systematics are combined to obtain one varied track collection which is then used to obtain one global vertex systematic and the corresponding varied collection of vertices. It is then used to compute the systematic variations of the observable $N_{\text{vtx,sel}}$.

For both tracking and vertexing systematics, only a down variation is provided with the track-dropping method. The up variations can be obtained by symmetrising the down variations observed in the analysis observables.

One way of visualising the effect of the systematics on the observable distributions is to compute for each systematic the difference in terms of number of events between the nominal distribution and the varied distribution. The differences can then be quadratically summed to obtain the total uncertainty envelop on the observable distributions. Fig. 7.44 shows the uncertainty envelop from the tracking and vertexing systematics for the track-based observables used in the analysis selection and for one signal sample with limited event selections.

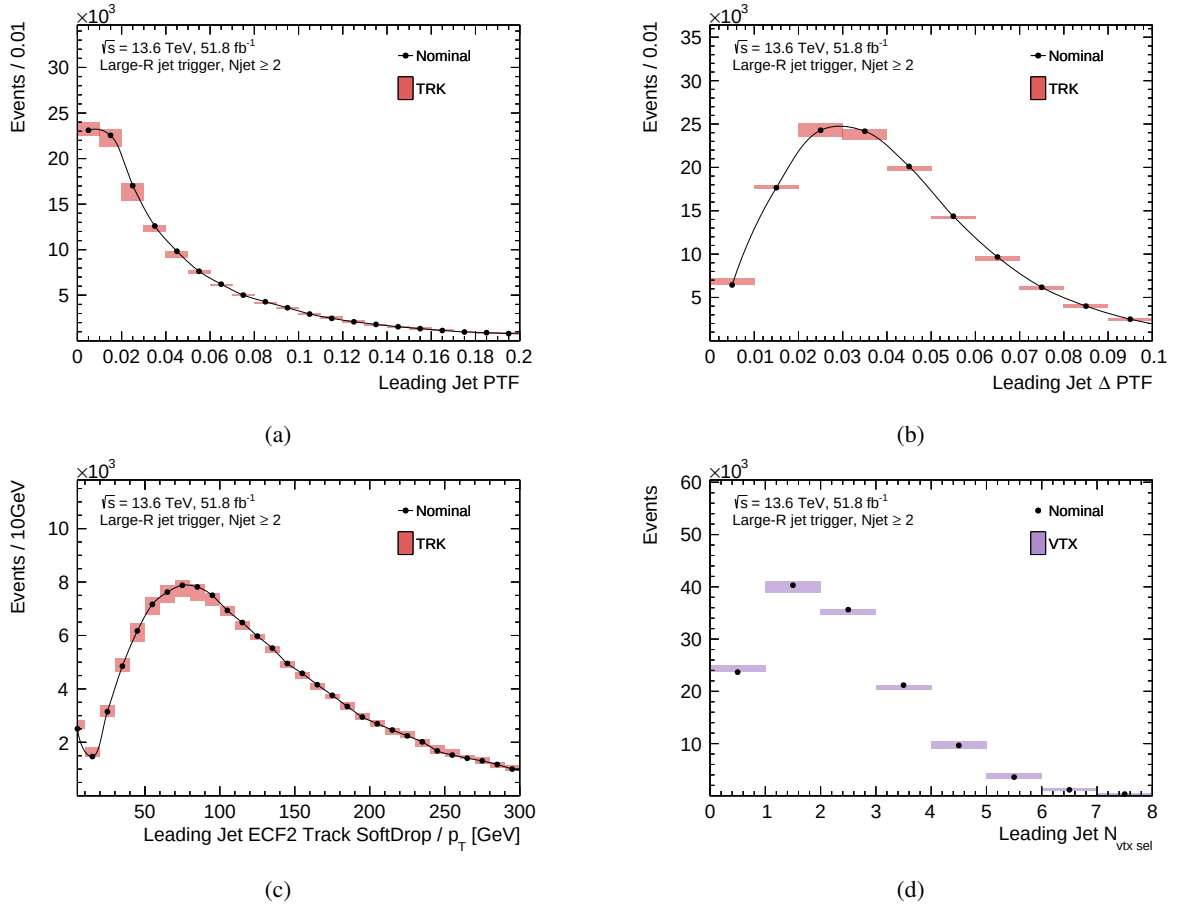


Figure 7.44: Systematic variations from tracking and vertexing systematics in the leading jet track-based observable distributions (PTF (a), ΔPTF (b), $ECF2/p_T$ (c) and $N_{vtx,sel}$ (d)) for the signal sample Ld20_rho40_pi10_Zp1500_15.

7.8.2 Jet systematics

Systematic uncertainties for jets are applied on the jet energy scale and/or mass scale to take into account the remaining uncertainties in the jet calibration. In principle, large- R jets being corrected in both energy and mass, systematic uncertainties in the mass scale need to be considered. However, as the large- R jet collection used in this analysis is based on the EMTopo small- R jets, only systematics affecting the small- R jets are used, such that only systematics in the energy scale are included which will then also propagate

to the mass computation. The main systematics in the energy scale of small- R jets are derived during the in-situ step of the jet calibration to take into account the relative precision of the applied corrections which can be affected by detector effects and the MC modelling. In addition, uncertainties in the jet flavours and jet composition in samples (quarks or gluon) are applied to take into account the difference in detector response between quark and gluon initiated jets and how both are modelled in MC simulations. Generally, quark initiated jets are well modelled by most generators, whereas the simulation of gluon jets can vary between several generators. It is thus necessary to consider the uncertainties in the jet flavours and the unknown jet composition of the analysis samples. These systematics are applied by varying the nominal value of the jet energy scale which leads to shifts in the jet kinematic distributions (e.g. p_T , m_{jj}). Additionally, a smearing of the jet energy is applied to take into account the uncertainty in the jet energy resolution which can differ between data and MC simulations. More information on this procedure can be found in [80].

The jet calibration is applied to SM QCD jets from the MC background samples but also to the EJs. To ensure that the jet calibration applies similarly to dark QCD jets than to SM QCD jets, the jet energy response distribution after calibration is inspected in multiple signal samples. The mode of the distributions are found to be within 5% of one (i.e. 5% away from closure) which is tolerable for large- R jets. Moreover, for MC background samples, the mode of the jet energy response distribution is observed to be slightly closer to one. The differences with the EJs are within the systematic uncertainties that are already applied to the SM QCD jets. No additional systematics thus deemed necessary for the calibration of EJs.

As in the previous tracking systematics plots, Fig. 7.45 shows the uncertainty envelop from jet systematics on the two main observables depending on the jet energy scale: the jet p_T and the 2-jet invariant mass.

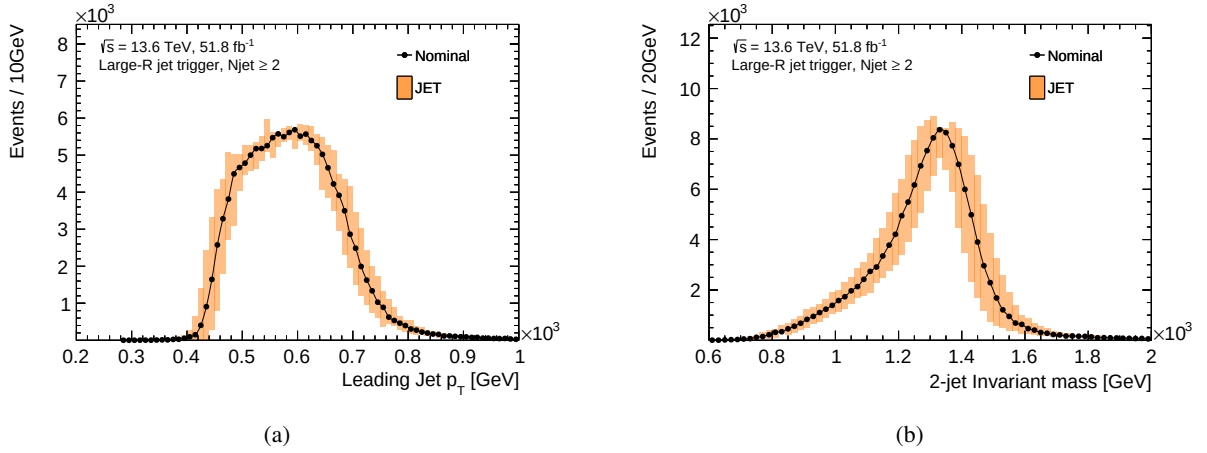


Figure 7.45: Systematic variations from jet systematics in the leading jet p_T (a) and the 2-jet invariant mass (b) distributions for the signal sample Ld20_rho40_pi10_Zp1500_15.

7.8.3 Generator systematics

Theoretical uncertainties in the SM parameters and the parton distribution function (PDF) of the protons need to be taken into account in MC simulations as they lead to variations in the structure of the events and on the measurement of the simulated process cross-section.

For the signal samples, the systematics varying the QCD coupling and scales are included in the form of variations in the amount of initial and final state radiations (ISR-FSR) in the events. These are the systematics that are the most impactful on the jet observables as variations in the number of additional radiation can alter the jet clustering and/or the jet substructure. These systematics, to avoid over-consumption of CPU time, are not obtained by re-generating the signal samples for each variation of the QCD parameters but instead are propagated to variations of the MC weights associated to each event through a re-weighting scheme similar to the pileup re-weighting procedure. Each variation of the QCD parameters leads to shifts of the nominal distribution of ISR-FSR (e.g. amount of radiation as a function of p_T) and the MC weights are modified accordingly so that the corresponding generated weighted event distribution is similarly shifted. This procedure is internal to the MC generator PYTHIA. Several variations are considered and only the systematics leading to the greatest variations of the signal yields are included.

No PDF systematics are included as they were not available yet for the signal samples at the time this thesis was written. It is not considered to be problematic as the main theoretical variations are expected to come from variations of the SM parameters impacting the jet structure and not from the PDF systematic that should mostly vary the global cross-section measurement. Moreover, it will be shown in the next section that the systematics on the signal yields have globally a small impact on the statistical results.

Additionally, no systematics on the dark shower dynamics are considered due to the general lack of knowledge of the actual phenomenology of the dark sector, such that it seems unnecessary to apply uncertainties on an unknown dynamic. Variations of the dark sector dynamics are, in fact, already taken into account through the different signal samples which vary the dark sector parameters.

The effect of the theoretical systematics on two observable distributions is shown in Fig. 7.46. The impact on the distributions is not as smooth as for the previous groups of systematics. This is possibly due to the way the uncertainty envelop is obtained. For each bin, only the systematic uncertainty leading to the maximum bin variation is used. Such a method can decorrelate the variations between bins and blur the global distribution variations.

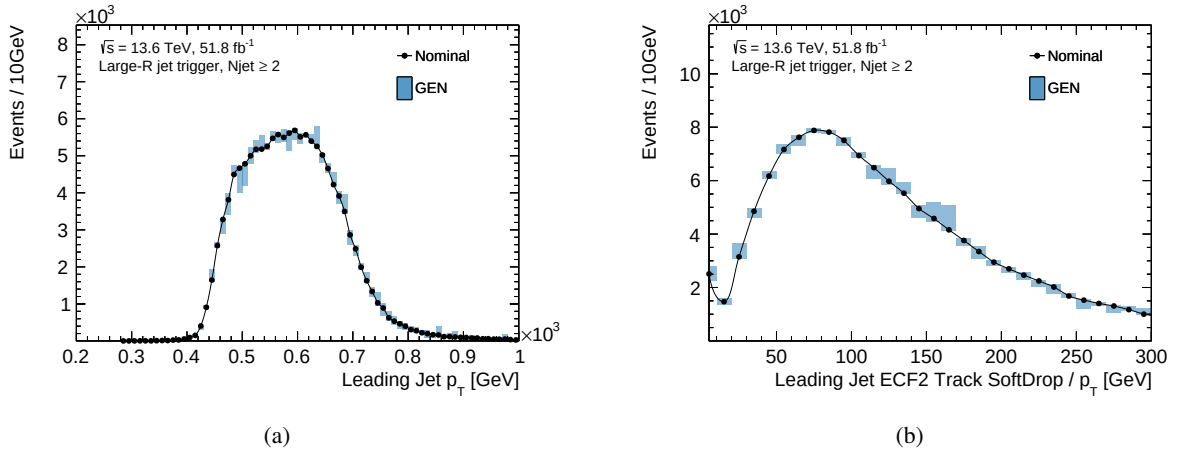


Figure 7.46: Systematic variations from theoretical systematics in the leading jet p_T and $ECF2/p_T$ distributions for the signal sample Ld20_rho40_pi10_Zp1500_15.

7.8.4 Pileup re-weighting systematic

Differences in the measurement of $\langle\mu\rangle$ between data and MC lead to uncertainties in the pileup re-weighting scheme applied to MC samples. In order to be taken into account, a systematic in the pileup weights is derived by shifting the $\langle\mu\rangle$ distribution in data by a constant scale factor and computing again the pileup weights corresponding to this new distribution and the unchanged $\langle\mu\rangle$ distribution of the MC sample. Two scale factors are used, leading to a down and up variations, and the obtained varied pileup weights are used as the systematic variations.

Fig. 7.47 shows the effect of the pileup systematic on two observable distributions. These two observables (ΔPTF and $ECF2/p_T$) are the most impacted by this systematic as these observables are dependent on the level of pileup in the event, thus shifts in the $\langle\mu\rangle$ distribution lead to global variations of the observable distribution.

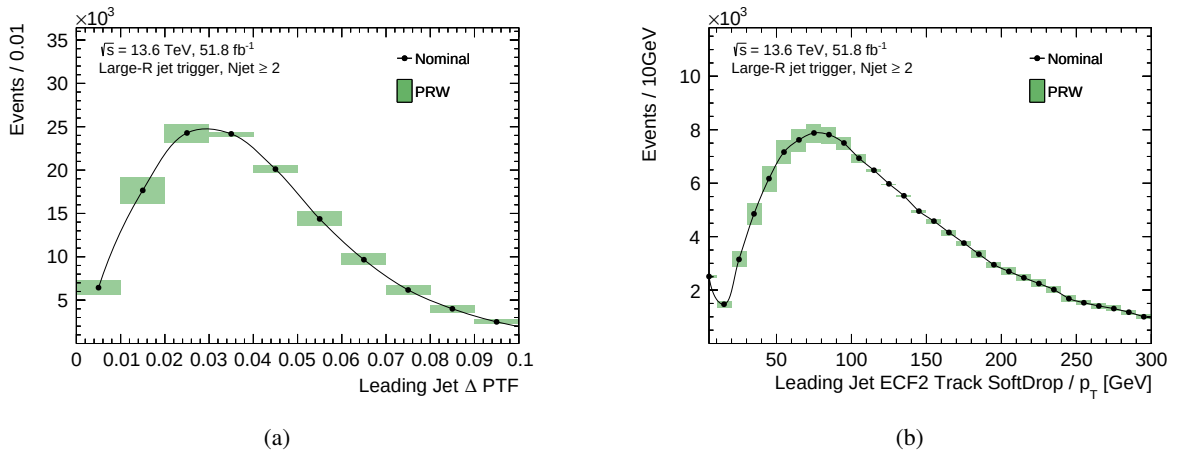


Figure 7.47: Systematic variations from pileup re-weighting systematic in the leading jet ΔPTF and $ECF2/p_T$ distributions for the signal sample Ld20_rho40_pi10_Zp1500_15.

7.8.5 Total systematic uncertainties in the signal yields in the SR and CRs

As firstly described above, each systematic is propagated through the analysis selections up to the ABCD plane and SR such that for each systematic a varied signal yield is obtained in each of the ABCD regions. The systematic uncertainty in the signal yield is taken as the difference between the varied and nominal yields. For each ABCD region, the negative yield uncertainties are quadratically summed to obtain the total down systematic uncertainty for the corresponding signal yield and vice versa for the up variations. The maximum between the up and down systematic uncertainties is finally taken as the total systematic uncertainty on each signal yield. The resulting uncertainties are presented in the form of the total relative uncertainty in the SR yield $\sigma^{SR} \equiv \sigma^A$ and the maximum up and down variations in the signal leakages $q_X = N_{\text{sig}}^X / N_{\text{sig}}^A$ obtained by propagation of the per-region yield uncertainties:

$$q_X^{\text{up}} = \frac{N_{\text{sig}}^X (1 + \sigma^X)}{N_{\text{sig}}^A (1 - \sigma^A)}$$

$$q_X^{\text{down}} = \frac{N_{\text{sig}}^X (1 - \sigma^X)}{N_{\text{sig}}^A (1 + \sigma^A)}$$

where σ^X is the relative total uncertainty on the signal yield in the region X .

Tab. 7.12 presents the obtained systematic uncertainties for each signal sample⁸. In addition, Tab. 7.13 shows the relative systematic uncertainty in the SR yield derived from each systematic group for all signal samples. The main observed trend is that the signal models with the highest $m_{Z'}$ show lower total systematic uncertainties in the SR yield. This is mainly correlated to the lower effect of the jet systematics as shifts in the jet p_T or m_{jj} distributions have a small impact on the signal acceptance due to the bulk of the distributions being far away from the selection thresholds. Similarly, the constantly lower (higher) uncertainties for the signal models with $c\tau_{\pi_D} = 50$ mm from the tracking (vertexing) systematics is a threshold effect from the selection on PTF ($N_{\text{vtx,sel}}$).

The systematics given in Tab. 7.12 are used in the statistical analysis in the next section.

Table 7.12: Total relative systematic uncertainties in the signal yield in the SR in percent and up and down systematic uncertainties in the leakages $q_X = N_{\text{sig}}^X / N_{\text{sig}}^A$ in percent.

Signal model	σ^{SR}	$q_B^{\text{up down}}$	$q_C^{\text{up down}}$	$q_D^{\text{up down}}$
Ld10_rho20_pi5_Zp1500_l5	± 13.7	$6.36^{+8.63}_{-5.03}$	$4.42^{+6.05}_{-3.41}$	$0.25^{+0.43}_{-0.18}$
Ld10_rho20_pi5_Zp1500_l50	± 15.9	$2.49^{+4.77}_{-1.78}$	$2.37^{+3.81}_{-1.56}$	$0.02^{+0.03}_{-0.01}$
Ld10_rho20_pi5_Zp3000_l5	± 8.23	$3.33^{+4.34}_{-2.82}$	$2.39^{+3.29}_{-1.94}$	$0.08^{+0.17}_{-0.05}$
Ld10_rho20_pi5_Zp3000_l50	± 7.69	$1.50^{+1.85}_{-1.03}$	$1.02^{+1.33}_{-0.61}$	$0.02^{+0.04}_{-0.01}$
Ld20_rho40_pi10_Zp1500_l5	± 13.7	$6.91^{+9.54}_{-5.44}$	$4.97^{+6.92}_{-3.95}$	$0.35^{+0.86}_{-0.22}$
Ld20_rho40_pi10_Zp1500_l50	± 12.7	$2.74^{+3.72}_{-1.87}$	$1.89^{+2.77}_{-1.34}$	$0.03^{+0.06}_{-0.00}$
Ld20_rho40_pi10_Zp3000_l5	± 7.64	$4.11^{+5.16}_{-3.26}$	$3.20^{+5.85}_{-2.55}$	$0.13^{+0.22}_{-0.07}$
Ld20_rho40_pi10_Zp3000_l50	± 6.32	$1.50^{+1.82}_{-1.21}$	$1.01^{+1.26}_{-0.72}$	$0.02^{+0.04}_{-0.00}$
Ld40_rho80_pi20_Zp1500_l5	± 13.2	$8.21^{+10.41}_{-6.73}$	$5.78^{+7.33}_{-4.63}$	$0.46^{+0.70}_{-0.25}$
Ld40_rho80_pi20_Zp1500_l50	± 11.6	$3.08^{+4.17}_{-2.13}$	$1.82^{+2.52}_{-1.36}$	$0.08^{+0.21}_{-0.04}$
Ld40_rho80_pi20_Zp3000_l5	± 8.18	$5.28^{+6.48}_{-4.24}$	$4.09^{+5.52}_{-3.44}$	$0.21^{+0.32}_{-0.14}$
Ld40_rho80_pi20_Zp3000_l50	± 5.45	$1.58^{+2.29}_{-1.37}$	$1.03^{+1.27}_{-0.86}$	$0.01^{+0.02}_{-0.00}$

⁸ Slight discrepancies in the values of q_X between Tab. 7.10 and Tab. 7.12 are observed and originate from the use of two sets of processed signal datasets which exhibit minor differences in objects reconstruction (jets, tracks, vertices).

Table 7.13: Relative systematic uncertainties in the signal yields in the SR for each systematic group in percent.

Signal model	Generator	Jet	Pileup re-weighting	Track	Vertex
Ld10_rho20_pi5_Zp1500_l5	± 1.38	± 10.8	± 0.50	± 7.79	± 2.93
Ld10_rho20_pi5_Zp1500_l50	± 4.27	± 13.3	± 1.94	± 3.73	± 6.36
Ld10_rho20_pi5_Zp3000_l5	± 3.38	± 1.06	± 0.59	± 7.02	± 2.36
Ld10_rho20_pi5_Zp3000_l50	± 7.20	± 1.12	± 2.16	± 4.57	± 5.60
Ld20_rho40_pi10_Zp1500_l5	± 1.78	± 11.1	± 0.35	± 7.75	± 2.05
Ld20_rho40_pi10_Zp1500_l50	± 1.25	± 11.1	± 1.23	± 3.45	± 4.84
Ld20_rho40_pi10_Zp3000_l5	± 5.42	± 1.00	± 0.25	± 7.40	± 1.53
Ld20_rho40_pi10_Zp3000_l50	± 3.55	± 0.88	± 1.42	± 4.34	± 4.12
Ld40_rho80_pi20_Zp1500_l5	± 1.62	± 10.9	± 0.10	± 7.09	± 1.67
Ld40_rho80_pi20_Zp1500_l50	± 1.72	± 10.0	± 1.24	± 3.29	± 4.31
Ld40_rho80_pi20_Zp3000_l5	± 3.15	± 1.11	± 0.25	± 7.97	± 1.31
Ld40_rho80_pi20_Zp3000_l50	± 1.42	± 0.75	± 1.08	± 3.97	± 3.50

7.9 Statistical analysis and results

The analysis strategy discussed in Sec. 7.4 is established with the definition of the SR, discussed in Sec. 7.6, and the data-driven background estimation method in place and validated in the MC background samples and most importantly in data, as discussed in Sec. 7.7. The development of the analysis is thus finalised. The final fundamental brick of the analysis is the statistical framework. This framework is needed to interpret the results that will be observed in data and measure their significance either to claim a potential discovery or to set upper limits on the signal cross-section. In particle physics, the statistical interpretation for BSM searches is generally built on likelihood-based tests of new physics [140]. For this analysis, the statistical framework is based on the ABCD-method likelihood that was briefly introduced in a simplified version in Sec. 7.4. The complete likelihood will be detailed here in the context of likelihood-based tests of new physics. This will be followed by the description of the formalism for statistical tests. A technical implementation will then be discussed. Finally, the expected results computed from the MC background samples will be presented.

7.9.1 Likelihood-based tests of new physics

As discussed in Sec. 7.4, the ABCD background estimation method can be interpreted as a four-bin model where each bin corresponds to the number of events in each of the ABCD regions. For each bin, the probability to observe N_{data}^X in data knowing that N_{evt}^X are expected is given by a Poisson distribution:

$$\text{Pois}(N_{\text{data}}^X | N_{\text{evt}}^X) = \frac{N_{\text{evt}}^{N_{\text{data}}^X} e^{-N_{\text{evt}}^X}}{N_{\text{data}}^X!} \quad (7.17)$$

In the analysis, it is the observed data N_{data}^X that is known and fixed and the expectation N_{evt}^X that is free and need to be constrained by the observation. This probability is then interpreted as a function of N_{evt}^X called the likelihood $\mathcal{L}(N_{\text{data}}^X | N_{\text{evt}}^X)$ which measures the likelihood of a certain N_{evt}^X to model well the observation N_{data}^X .

The ABCD plane consists of four regions $\{A, B, C, D\}$ corresponding to four separate observations N_{data}^X that are supposed to be independent such that the total likelihood for observing **data** = $\{N_{\text{data}}^A, N_{\text{data}}^B, N_{\text{data}}^C, N_{\text{data}}^D\}$ is given by the product of the likelihood of the four regions:

$$\begin{aligned} \mathcal{L}_{\text{tot}}(\mathbf{data} | \{N_{\text{evt}}^A, N_{\text{evt}}^B, N_{\text{evt}}^C, N_{\text{evt}}^D\}) &= \prod_{X \in \{A, B, C, D\}} \mathcal{L}(N_{\text{data}}^X | N_{\text{evt}}^X) \\ &= \prod_{X \in \{A, B, C, D\}} \text{Pois}(N_{\text{data}}^X | N_{\text{evt}}^X) \end{aligned} \quad (7.18)$$

As discussed in Sec. 7.4, the expected numbers of events in each region are parametrised in terms of the signal and background contributions:

$$\begin{aligned}
N_{\text{evt}}^A &= \mu + \mu_B \\
N_{\text{evt}}^B &= \mu q_B + \mu_B \tau_B \\
N_{\text{evt}}^C &= \mu q_C + \mu_B \tau_C \\
N_{\text{evt}}^D &= \mu q_D + \mu_B \tau_B \tau_C
\end{aligned}$$

μ represents the number of signal events in the SR, it is the only parameter of interest for the analysis, and is referred to as the signal strength⁹. μ_B represents the number of background events in the SR and (τ_B, τ_C) are the parameters enforcing the ABCD relationship for the background events. These three parameters are nuisance parameters as they are not used for the physical interpretation but are needed for the statistical interpretation, to constrain the model and the estimation of μ . The signal leakages $q_X = N_{\text{sig}}^X / N_{\text{sig}}^A$ are fixed parameters that are computed for each signal model from the MC simulations and supplied when constructing the likelihood. Additional nuisance parameters corresponding to the systematic uncertainties in the signal yields in the ABCD regions, σ^X , are considered. Each systematic is included in the likelihood with a constraint term $C(N_{\text{sig}}^X, \sigma^X)$. Generally, this constraint term is taken as a gaussian distribution centred on a nominal value, $N_{\text{sig,exp}}$, assumed in the data and with a standard deviation taken as the systematic uncertainty:

$$C(N_{\text{sig}}^X, \sigma^X) = \mathcal{G}(N_{\text{sig,exp}}^X | N_{\text{sig}}^X, \sigma_X) \quad (7.19)$$

In the following $\theta = \{\mu_B, \tau_B, \tau_C, \sigma^A, \sigma^B, \sigma^C, \sigma^D\}$ will denote the nuisance parameters in the likelihood. The complete likelihood for observing **data** is then given by:

$$\begin{aligned}
\mathcal{L}_{\text{tot}}(\mathbf{data} | \mu, \theta) = & \text{Pois}(N_{\text{data}}^A | \mu + \mu_B) \\
& \times \text{Pois}(N_{\text{data}}^B | \mu q_B + \mu_B \tau_B) \\
& \times \text{Pois}(N_{\text{data}}^C | \mu q_C + \mu_B \tau_C) \\
& \times \text{Pois}(N_{\text{data}}^D | \mu q_D + \mu_B \tau_B \tau_C) \\
& \times C(\mu, \sigma^A) C(\mu q_B, \sigma^B) C(\mu q_C, \sigma^C) C(\mu q_D, \sigma^D)
\end{aligned} \quad (7.20)$$

This total likelihood can be used to constrain μ and the nuisance parameters θ to best fit the observations by finding the maximum of $\mathcal{L}_{\text{tot}}$. The values of the parameters maximising the likelihood are denoted below with a " $\hat{}$ " and referred to as the estimators of the parameters. μ can also be fixed to 0 to constrain the nuisance parameters in a background-only hypothesis which is then equivalent to simply applying the standard ABCD background estimation method.

⁹ The term signal strength is used here in analogy with the usual statistical interpretation nomenclature found in the literature in which the signal strength μ is used as a scaling factor to the expected number of signal events in a region ($N_{\text{sig}} = \mu \times N_{\text{sig,exp}}$) and is defined as the ratio of the observed and theoretical cross-sections ($\mu = \sigma_{\text{observed}} / \sigma_{\text{theory}}$). However, in this analysis μ is taken directly as the number of signal events in the SR and is used to scale the number of signal events in the other B, C and D regions.

However, finding the parameters maximising the likelihood is not enough to conclude on the results of the analysis, there is a need for a consistent formalism for statistical tests providing a measurement of the significance of the observed results. In the context of particle physics and the frequentist approach, the statistical test consists in measuring the agreement of the observed data with a given hypothesis and comparing the agreements obtained for different hypotheses. For discovery of a new signal process, like in this analysis, two hypotheses are generally tested against. First, the null or background-only hypothesis, referred to as H_0 , supposes only a contribution from the SM background in the SR and thus corresponds to $\mu = 0$. Second, the alternative signal + background hypothesis, referred to as H_1 , where a signal is expected in the SR with a free signal strength.

The agreement of the observed data with a given hypothesis H is measured through the p -value. The p -value is the probability, under the assumption that H is true, to find observations of equal or greater incompatibility with the predictions of H . The hypothesis H is considered excluded or rejected if the p -value is measured under a specific threshold α usually taken as 0.05 for excluding signal hypotheses with a 95% confidence level (CL). α can be interpreted as the probability of rejecting H even though the hypothesis is true.

In a simplified approach, the incompatibility between the observations and H can be measured simply through the number of events found in the SR compared to the prediction of H , knowing that the underlying distribution is a Poisson distribution. In real analyses, the p -value is generally measured through what is called a test statistic.

7.9.1.1 Test statistics

In the LHC experiments, the basic test statistic used to test a hypothesised value of μ (i.e. a hypothesis H) is the profile likelihood ratio $\lambda(\mu)$ defined as:

$$\lambda(\mu) = \frac{\mathcal{L}_{\text{tot}}(\mu, \hat{\hat{\theta}}(\mu))}{\mathcal{L}_{\text{tot}}(\hat{\mu}, \hat{\theta})} \quad (7.21)$$

The denominator is the unconditionally maximised likelihood where $\hat{\mu}$ and $\hat{\theta}$ are the estimators of the parameters corresponding to the maximum likelihood. The numerator is the conditional maximised likelihood for the tested μ , where $\hat{\hat{\theta}}$ corresponds to the value of θ maximising the likelihood for this specific μ . It is a function of the signal strength outputting a simple number between 0 and 1 which allows to characterise the agreement of the data with several hypothesised μ , with λ close to 1 implying a good agreement.

It is then convenient to define t_μ as the actual test statistic:

$$t_\mu = -2 \ln \lambda(\mu) \quad (7.22)$$

such that a greater value of t_μ corresponds to higher level of incompatibility between the data and the tested signal strength. It is thus a function that can be used to obtain an analytical definition of the p -value:

$$p_\mu = \int_{t_{\mu, \text{obs}}}^{\infty} f(t_\mu | \mu') dt_\mu \quad (7.23)$$

where $f(t_\mu|\mu')$ represents the probability density function of t_μ under the assumption μ' and $t_{\mu,\text{obs}}$ is the observed value of the test statistic in data for this specific μ . In $f(t_\mu|\mu')$, the subscript of t corresponds to the signal strength that is being tested in the hypothesis H and μ' is the value of the signal strength that is assumed in the data. For this definition of p_μ both are taken equal ($\mu' = \mu$) as the purpose is to test a hypothesis H assuming it is true in data but they can be differentiated to test and exclude different hypotheses.

Two specific variations of the test statistic are used in the context of search of positive new signal processes, one for discovery purposes and the other for setting upper limits on the signal strength.

Test statistics for discovery of a positive signal When searching for the presence of signal events in a certain region of phase space, it is expected that the signal contribution can only lead to an increase of events, in case no interference is expected, meaning that $\mu \geq 0$. For any model with a positive signal contribution, if the data favours $\hat{\mu} < 0$ then the best agreement (i.e. $\lambda = 1$) between data and any hypothesised positive μ will be obtained for $\mu = 0$. In this particular case, a new definition of λ can be made:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{\mathcal{L}_{\text{tot}}(\mu, \hat{\theta}(\mu))}{\mathcal{L}_{\text{tot}}(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0, \\ \frac{\mathcal{L}_{\text{tot}}(\mu, \hat{\theta}(\mu))}{\mathcal{L}_{\text{tot}}(0, \hat{\theta}(0))} & \hat{\mu} < 0 \end{cases} \quad (7.24)$$

Such a definition allows to avoid rejecting the background-only hypothesis in the case of over-estimation or fluctuations in the background contributions in data that could lead to $\hat{\mu} < 0$. This new definition can then be used to define a new test statistic $\tilde{t}_\mu$:

$$\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu) \quad (7.25)$$

In the context of the discovery of a positive signal, the null hypothesis corresponds to the background-only hypothesis with $\mu = 0$ and one wants to test it against the alternative hypothesis which corresponds to the signal + background hypothesis with $\mu > 0$. By rejecting the null hypothesis, one can a priori claim the discovery of a new signal. The test statistic corresponding to the background-only hypothesis is $\tilde{t}_0$, that is specifically denoted q_0 :

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0 \end{cases} \quad (7.26)$$

With this definition based on $\tilde{\lambda}$ only a disagreement between data and the background-only hypothesis for $\hat{\mu} \geq 0$ can be used to exclude this hypothesis. If the data favours $\hat{\mu} < 0$ then $q_0 = 0$. However, q_0 will increase with $\hat{\mu}$, meaning that the disagreement between data and the background-only hypothesis will increase with the event yield increasing above the background expectation. This incompatibility is measured with the p -value that is now defined by:

$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0) dq_0 \quad (7.27)$$

where $f(q_0|0)$ is the probability density function of q_0 under the assumption of $\mu = 0$, the background-only hypothesis.

Test statistic for upper limits If no excess is observed, the null hypothesis becomes the signal + background hypothesis with $\mu > 0$ and one wants to exclude it against the alternative background-only hypothesis with $\mu = 0$. Excluding the signal + background hypothesis corresponds to finding the maximum signal strength μ_{up} for which any hypothesis with $\mu > \mu_{\text{up}}$ will be excluded at a certain confidence level $1 - \alpha$ (usually 95%). μ_{up} is referred to as the upper limit on the signal strength at 95% confidence level. In order to obtain such upper limits, the test statistic able to exclude the signal + background hypothesis is defined as $\tilde{q}_\mu$:

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu \end{cases} \quad (7.28)$$

As the point is to exclude hypothesised μ , meaning exclude μ for which the data shows greater incompatibility than with the data-favoured $\hat{\mu}$, only the hypotheses with $\hat{\mu} \leq \mu$ are considered. If the data fluctuates upward, above the tested μ , there is no point in testing such hypothesis for setting upper limits, in such cases the test statistic is set to 0. Additionally, $\tilde{\lambda}$ is used here as models with positive signal contributions are still considered.

The level of agreement between data and the hypothesised μ is again obtained with the p -value:

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu) dq_\mu \quad (7.29)$$

The observed upper limit on the signal strength in data is then determined by scanning the possible values of μ and computing for each the corresponding p_μ . The upper limit is then taken as the maximum μ for which $p_\mu > 0.05$ (i.e. for which the signal + background hypothesis cannot be rejected).

In some cases where the analysis is poorly sensitive to a model and more generally when the separation between the background-only ($\mu = 0$) and the signal + background ($\mu > 0$) hypotheses is very limited, it is possible that the upper limit based on p_μ can lead to exclusion of very small signal strength if there are some fluctuations in the background contributions. Such poorly sensitive analyses should not be able to exclude very low values of μ . To avoid such an effect, the upper limits are usually obtained by comparing a modified p -value to the threshold α , called the CL_s value and defined as:

$$\text{CL}_s = \frac{p_\mu}{1 - p_b} \quad (7.30)$$

where p_b is the probability to find data with greater compatibility with the background-only hypothesis:

$$p_b = \int_{-\infty}^{q_{\mu,\text{obs}}} f(\tilde{q}_\mu|0) dq_\mu \quad (7.31)$$

For an analysis with poor sensitivity, p_b will be large for any hypothesised μ leading to an increase in CL_s and thus the exclusion of small values of the signal strength will be avoided.

In addition, the expected upper limit $\mu_{\text{up,exp}}$ is usually also provided as a reference. It corresponds to the upper limit assuming the background-only hypothesis in data. It is computed based on the same definition of the p -value but with $\mu' = 0$:

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu|0) dq_\mu \quad (7.32)$$

$\mu_{\text{up,exp}}$ is taken as the maximum μ for which $p_\mu > 0.05$. It is often computed using MC background samples as the observed data, in which the background-only hypothesis and the absence of signal is evident.

7.9.1.2 Approximation of distribution of the profile likelihood ratio

In order to be able to use these different definitions of the p -values, it is necessary to know the sampling distribution of the test statistics. For discovery, one needs the distribution $f(q_0|0)$. For setting upper limits, one needs the distributions $f(\tilde{q}_\mu|\mu)$ and $f(\tilde{q}_\mu|0)$. More generally, one needs the distribution $f(t_\mu|\mu')$ where, as already discussed before, the subscript μ of t corresponds to the hypothesised signal strength and μ' is the signal strength assumed in data. These distributions can be sampled by generating MC pseudo-data generated under several hypothesis assumptions. This method is however CPU-intensive and generally not used. The second option is to use an approximation of the profile likelihood ratio valid for large data samples (which is usually the case in particle physics and especially in the LHC experiments). This approximation is due to a result from Wald [141] that shows that for a profile likelihood ratio with one parameter of interest (one degree of freedom) μ :

$$t_\mu = -2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + O(1/\sqrt{N}) \quad (7.33)$$

where $\hat{\mu}$ is assumed to follow a Gaussian distribution centred on μ' (assumption in data) with a standard deviation σ and N is simply the statistic of the data sample. By neglecting $O(1/\sqrt{N})$ (large sample limit) it is possible to show that t_μ follows a non-central chi-square distribution:

$$f(t_\mu; \Lambda) = \frac{1}{2\sqrt{t_\mu}} \frac{1}{\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} + \sqrt{\Lambda}\right)^2\right) + \exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} - \sqrt{\Lambda}\right)^2\right) \right] \quad (7.34)$$

where Λ is referred to as the non-centrality parameter defined as:

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2} \quad (7.35)$$

This approximation allows to greatly simplify the use of the p -values. However, there is still one missing information which is the standard deviation σ of $\hat{\mu}$. This quantity can be estimated using an artificial dataset called an "Asimov" dataset. It is defined such that when used to evaluate the estimators of the model parameters (μ and the nuisance parameters) the obtained estimators are equal to the true parameter values assumed in the real dataset. This leads to $\hat{\mu} = \mu'$ and $\hat{\theta} = \theta$. This dataset is generally constructed as

the dataset corresponding to the estimated signal and background distributions¹⁰. For the statistical model $\mathcal{L}_{\text{tot}}$ considered in this analysis, the "Asimov" dataset is then constructed as:

$$\begin{aligned} N_{\text{data,Asimov}}^A &= \mu' + \hat{\mu}_B \\ N_{\text{data,Asimov}}^B &= \mu' q_B + \hat{\mu}_B \hat{\tau}_B \\ N_{\text{data,Asimov}}^C &= \mu' q_C + \hat{\mu}_B \hat{\tau}_C \\ N_{\text{data,Asimov}}^D &= \mu' q_D + \hat{\mu}_B \hat{\tau}_B \hat{\tau}_C \end{aligned}$$

The "Asimov" dataset can be used to evaluate the likelihood $\mathcal{L}_A$ and obtain the corresponding profile likelihood ratio λ_A :

$$\lambda_A(\mu) = \frac{\mathcal{L}_A(\mu, \hat{\theta})}{\mathcal{L}_A(\hat{\mu}, \hat{\theta})} = \frac{\mathcal{L}_A(\mu, \hat{\theta})}{\mathcal{L}_A(\mu', \theta)} \quad (7.36)$$

The test statistic for the "Asimov" dataset is then approximated by:

$$t_{\mu,A} = -2 \ln \lambda_A(\mu) \approx \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda \quad (7.37)$$

This provides an estimation for the non-centrality parameter Λ and consequently for σ :

$$\sigma^2 = \frac{(\mu - \mu')^2}{t_{\mu,A}} \quad (7.38)$$

This parameter is affected by the nuisance parameters and in particular by the systematic uncertainties.

These approximations of the sampling distribution $f(t_\mu|\mu')$ can then be applied on more specific cases to obtain the approximated distributions $f(q_0|0)$, $f(\tilde{q}_\mu|\mu)$ and $f(\tilde{q}_\mu|0)$ and the corresponding p -values for any hypothesised μ . This method is the one used in the statistical framework that is exploited to interpret the analysis results.

7.9.2 Statistical framework

The statistical framework used for this analysis is based on *pyhf* [142, 143]. It is a pure-python module providing tools for constructing statistical models and implementing the likelihood interpretation formalism described in the previous subsection. The analysis uses the overlay *abcd-pyhf* [144] that implements the likelihood model corresponding to the ABCD method given by Eq. 7.20. Only one systematic is considered on the signal strength in this framework. This systematic is taken as the total systematic in the SR yield as it will be shown in section 7.9.3.3 that the systematics in the signal yields in the CRs lead to negligible variations in the obtained upper limits. The statistical model is thus reduced from 7.20 to:

¹⁰ It is as if the observed data was exactly mimicking what is estimated.

$$\begin{aligned}
\mathcal{L}_{\text{tot}}(\mathbf{data}|\mu, \theta) = & \text{Pois}(N_{\text{data}}^A | \mu + \mu_B) \\
& \times \text{Pois}(N_{\text{data}}^B | \mu q_B + \mu_B \tau_B) \\
& \times \text{Pois}(N_{\text{data}}^C | \mu q_B + \mu_B \tau_C) \\
& \times \text{Pois}(N_{\text{data}}^D | \mu q_D + \mu_B \tau_B \tau_C) \\
& \times C(\mu, \sigma^{SR})
\end{aligned} \tag{7.39}$$

For setting upper limits, the test statistic considered is $\tilde{q}_\mu$ defined in Eq. 7.28. The modified p -value CL_s defined in Eq. 7.30 is used for the observed limits and p_μ defined in Eq. 7.32 for the expected limits. The statistic model takes as input the observed data yield N_{data}^X in each of the four regions of the ABCD plane along, for each signal, the signal leakages q_X in the CRs.

7.9.3 Results and limits on the signal cross-section from MC simulations

As stated in the introduction, this analysis is still under approval in the ATLAS collaboration such that only blinded data are accessible. The blinded data allowed to validate the ABCD background estimation method but do not allow to actually obtain the final statistical interpretation of the analysis on the SR. Thus, this subsection will present the expected sensitivities computed from MC simulations only, meaning that only the expected signal cross-section upper limits in an absence of signal in the observed data can be obtained.

7.9.3.1 Nominal test of the statistical framework in MC simulations

A first check of the statistical model is to compute the background estimation in the background-only hypothesis where the data provided to the model is taken as the MC background dijet + top samples. The observed data are thus the same as in Sec. 7.7.1:

$$\begin{aligned}
N_{\text{data}}^A &= 27.2 \pm 5.2 \\
N_{\text{data}}^B &= 409.8 \pm 20.2 \\
N_{\text{data}}^C &= 352.3 \pm 18.8 \\
N_{\text{data}}^D &= 4974.4 \pm 70.5
\end{aligned} \tag{7.40}$$

In the background-only hypothesis, the signal strength is set to 0, $\mu = 0$, such that:

$$\begin{aligned}
\mathcal{L}_{\text{tot}}(\mathbf{data}|0, \theta) = & \text{Pois}(N_{\text{data}}^A | \mu_B) \\
& \times \text{Pois}(N_{\text{data}}^B | \mu_B \tau_B) \\
& \times \text{Pois}(N_{\text{data}}^C | \mu_B \tau_C) \\
& \times \text{Pois}(N_{\text{data}}^D | \mu_B \tau_B \tau_C) \\
& \times C(0, \sigma^{SR})
\end{aligned} \tag{7.41}$$

The total likelihood is then used to only constrain the nuisance parameters θ to obtain the estimators most suited to the observed data:

$$\begin{aligned}\hat{\mu}_B &= 29.0 \pm 2.1 \\ \hat{\tau}_B &= 14.1 \pm 0.8 \\ \hat{\tau}_C &= 12.1 \pm 0.6\end{aligned}\tag{7.42}$$

which correspond to:

$$\begin{aligned}N_{\text{bkg,estimated}}^A &= 29.0 \pm 2.1 \\ N_{\text{bkg,estimated}}^B &= 409.8 \pm 37.0 \\ N_{\text{bkg,estimated}}^C &= 352.3 \pm 30.8 \\ N_{\text{bkg,estimated}}^D &= 4974.4 \pm 513.2\end{aligned}\tag{7.43}$$

It can be observed that in the background-only hypothesis, the estimated numbers of background events in the CRs are fixed to the observed numbers and used to constrain the SR background yield μ_B . This test is thus completely analogue to the standard ABCD background estimation method described in Sec. 7.4 and previously used in Sec. 7.7 and lead to the same estimation of N_{bkg}^A .

A second nominal test is then performed by letting the signal strength μ free to ensure that the statistical framework estimate $\hat{\mu} = 0$. This is a test of the signal + background hypothesis. The total likelihood is thus the one given in Eq. 7.20. In addition to the observed data (from the MC background samples), the signal leakages q_X and SR yield systematic uncertainty are provided for each tested signal model. Only the results for the signal model Ld20_rho40_pi10_Zp1500_15 is shown here as all models lead to similar results. For this model, in percent:

$$\begin{aligned}q_B &= 6.78 \\ q_C &= 4.97 \\ q_D &= 0.35 \\ \sigma_{SR} &= 13.7\end{aligned}\tag{7.44}$$

By constraining $\mathcal{L}_{\text{tot}}$ one obtain:

$$\begin{aligned}\hat{\mu}_B &= 28.8 \pm 2.0 \\ \hat{\tau}_B &= 14.2 \pm 0.8 \\ \hat{\tau}_C &= 12.2 \pm 0.6 \\ \hat{\mu} &= 0.0 \pm 9.0\end{aligned}\tag{7.45}$$

As expected the statistical model estimates $\hat{\mu} = 0$. In addition, the signal leakages slightly alter the background estimation.

These two nominal checks allow to ensure that the statistical model and framework behave well with the MC background simulations by providing a consistent background estimation and estimating that there is no signal in the background samples.

A sanity check can also be performed by artificially injecting a signal contribution in the SR observed yield (N_{data}^A) and check if the signal strength estimation is able to recover it. With the same setup as the signal + background test, 10 events are added artificially in the observed data in the SR by simply changing $N_{\text{data}}^A = 27.2$ to $N_{\text{data}}^A = 37.2$. The obtained estimations are:

$$\begin{aligned}\hat{\mu}_B &= 28.9 \pm 2.2 \\ \hat{\tau}_B &= 14.1 \pm 0.8 \\ \hat{\tau}_C &= 12.2 \pm 0.6 \\ \hat{\mu} &= 8.30 \pm 6.53\end{aligned}\tag{7.46}$$

The statistical framework shows to be able to recover a large part of the signal injection without altering the background estimation.

7.9.3.2 Predicted upper limits on the signal cross-section

In the absence of a signal, upper limits on the signal strength μ_{up} can be set, both expected in the background-only hypothesis and observed in the signal + background hypothesis. As the unblinded data are not accessible, the limits can only be computed from the MC backgrounds samples and only the expected upper limits will be presented as there is not much interest in computing the observed limits from MC samples with an absence of signal. It is also usual to quote the error band on the nominal expected upper limit with $\pm 1\sigma$ and $\pm 2\sigma$ variations where σ is the standard deviation of $\hat{\mu}$ as discussed in Sec. 7.9.1.2 and computed with Eq. 7.38. The statistical framework thus provides the limits $\mu_{\text{up,exp}}$ and:

$$\begin{aligned}\mu_{\text{up,exp}}^{\pm 1\sigma} &= \mu_{\text{up,exp}} \pm 1\sigma \\ \mu_{\text{up,exp}}^{\pm 2\sigma} &= \mu_{\text{up,exp}} \pm 2\sigma\end{aligned}\tag{7.47}$$

These upper limits are computed using $\mathcal{L}_{\text{tot}}$ and taking the MC background samples as the observed data. It allows to provide predictions on the limits than can be expected in the real data in the absence of signal. Tab. 7.14 gives the obtained expected limits on μ obtained from the MC background samples for each signal model.

The upper limits on the signal strength are then converted to upper limits on the signal cross-section. The number of signal events in the SR, $N_{\text{sig}}^{\text{SR}}$, can be expressed as:

$$N_{\text{sig}}^{\text{SR}} = \mathcal{A}_{\text{tot}} \times \sigma_{\text{sig}} k f_{\text{eff}} \times \mathcal{L}_{\text{int}}\tag{7.48}$$

where $\mathcal{A}_{\text{tot}}$ is the total signal acceptance in the SR and σ_{sig} is the signal cross-section.

Table 7.14: Predicted upper limits on the signal strength μ obtained in MC simulations. The total signal acceptance in percent is reminded in the second column and the SR yield uncertainty in percent in the third column.

Signal model	$\mathcal{A}_{\text{tot}}$	σ^{SR}	Expected limits in the signal strength				
			$\mu_{\text{up,exp}}^{-2\sigma}$	$\mu_{\text{up,exp}}^{-1\sigma}$	$\mu_{\text{up,exp}}$	$\mu_{\text{up,exp}}^{+1\sigma}$	$\mu_{\text{up,exp}}^{+2\sigma}$
Ld10_rho20_pi5_Zp1500_15	18.0	13.7	6.57	8.97	13.01	19.44	28.54
Ld10_rho20_pi5_Zp1500_150	7.3	15.9	6.56	8.99	13.15	19.82	29.56
Ld10_rho20_pi5_Zp3000_15	39.2	8.23	6.50	8.85	12.69	18.53	26.44
Ld10_rho20_pi5_Zp3000_150	19.2	7.69	6.48	8.83	12.65	18.44	26.24
Ld20_rho40_pi10_Zp1500_15	24.2	13.7	6.57	8.98	13.03	19.46	28.57
Ld20_rho40_pi10_Zp1500_150	14.1	12.7	6.53	8.92	12.90	19.15	27.95
Ld20_rho40_pi10_Zp3000_15	49.0	7.64	6.51	8.86	12.69	18.49	26.32
Ld20_rho40_pi10_Zp3000_150	31.5	6.32	6.48	8.81	12.60	18.31	25.92
Ld40_rho80_pi20_Zp1500_15	23.9	13.2	6.58	8.98	13.01	19.39	28.38
Ld40_rho80_pi20_Zp1500_150	15.5	11.6	6.52	8.90	12.84	18.96	27.52
Ld40_rho80_pi20_Zp3000_15	48.0	8.18	6.52	8.87	12.73	18.58	26.51
Ld40_rho80_pi20_Zp3000_150	32.3	5.45	6.47	8.81	12.58	18.24	25.77

Thus, the upper limit on the signal cross-section σ_{up} is given by:

$$\sigma_{\text{up}} = \frac{\mu_{\text{up}}}{\mathcal{A} \times k f_{\text{eff}} \times \mathcal{L}_{\text{int}}} \quad (7.49)$$

Using $\mu_{\text{up,exp}}$ and $\mu_{\text{up,exp}}^{\pm 1\sigma}$, $\mu_{\text{up,exp}}^{\pm 2\sigma}$, the expected band upper limits on the signal cross-section are obtained: $\sigma_{\text{up,exp}}$, $\sigma_{\text{up,exp}}^{\pm 1\sigma}$, $\sigma_{\text{up,exp}}^{\pm 2\sigma}$.

The total signal acceptance $\mathcal{A}_{\text{tot}}$ were given in Tab. 7.10 and the total integrated luminosity of the early Run 3 data taking period (2022+2023) amount for 51.8 fb^{-1} . The predicted upper limits on the signal cross-section in fb for each signal model are provided in Tab. 7.15.

The upper limits are also reported in Fig. 7.48. The nominal expected limits are closely related to the signal acceptances and the width of the expected band correlated to the systematic uncertainties.

In the real data, when available, if the observed limit for a model is found to be above and outside the expected band, it could constitute evidences for a new signal in data. In any cases, all signal cross-sections above the observed limits would be ruled out. In contrast, the cross-sections lower than the observed limits can not be rejected as the analysis would not be sensitive enough to be able to observe events in data. Improvement on the signal cross-section constraints can be obtained by either increasing the integrated luminosity (more data) or increasing the signal acceptance in the SR while maintaining the same background rejection.

Table 7.15: Predicted upper limits on the signal cross-section in fb obtained in MC simulations. The total signal acceptance in percent is reminded in the second column and the SR yield uncertainty in percent in the third column.

Signal model	$\mathcal{A}_{\text{tot}}$	σ^{SR}	Expected limits [fb]				
			$\sigma_{\text{up,exp}}^{-2\sigma}$	$\sigma_{\text{up,exp}}^{-1\sigma}$	$\sigma_{\text{up,exp}}$	$\sigma_{\text{up,exp}}^{+1\sigma}$	$\sigma_{\text{up,exp}}^{+2\sigma}$
Ld10_rho20_pi5_Zp1500_15	18.0	13.7	0.96	1.31	1.90	2.84	4.17
Ld10_rho20_pi5_Zp1500_150	7.3	15.9	2.35	3.21	4.70	7.08	10.6
Ld10_rho20_pi5_Zp3000_15	39.2	8.23	0.70	0.95	1.36	1.99	2.84
Ld10_rho20_pi5_Zp3000_150	19.2	7.69	1.42	1.93	2.76	4.03	5.73
Ld20_rho40_pi10_Zp1500_15	24.2	13.7	0.74	1.01	1.47	2.19	3.21
Ld20_rho40_pi10_Zp1500_150	14.1	12.7	1.26	1.72	2.49	3.70	5.40
Ld20_rho40_pi10_Zp3000_15	49.0	7.64	0.57	0.78	1.11	1.62	2.31
Ld20_rho40_pi10_Zp3000_150	31.5	6.32	0.88	1.20	1.72	2.49	3.53
Ld40_rho80_pi20_Zp1500_15	23.9	13.2	0.75	1.02	1.48	2.21	3.23
Ld40_rho80_pi20_Zp1500_150	15.5	11.6	1.14	1.56	2.25	3.33	4.83
Ld40_rho80_pi20_Zp3000_15	48.0	8.18	0.55	0.75	1.07	1.57	2.24
Ld40_rho80_pi20_Zp3000_150	32.3	5.45	0.81	1.11	1.58	2.29	3.23

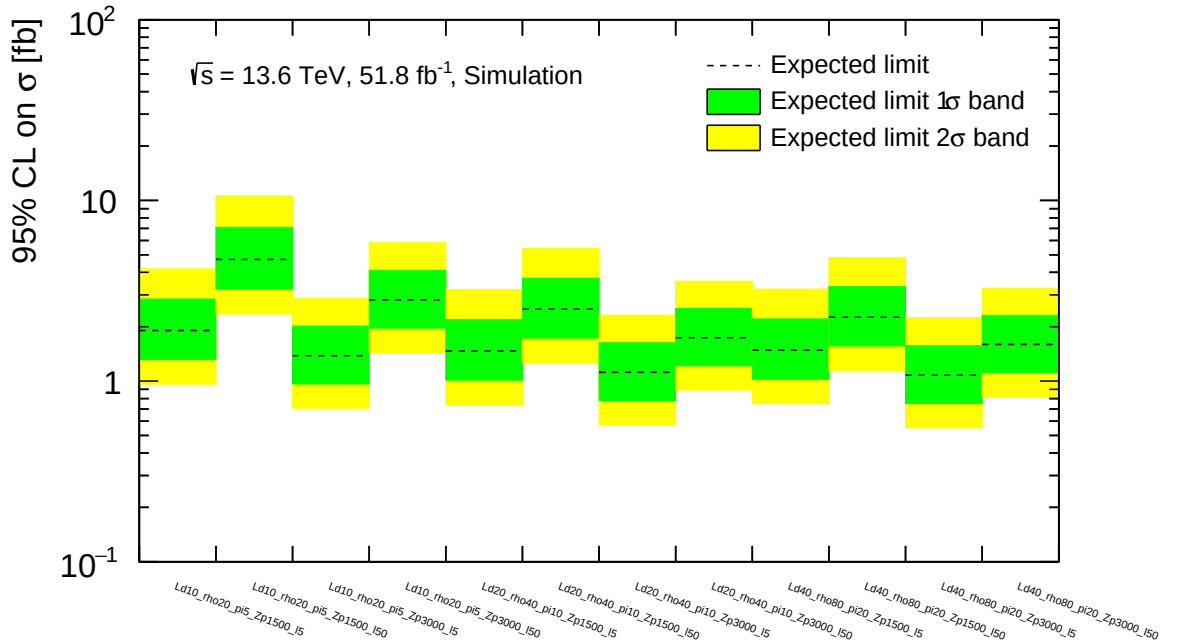


Figure 7.48: Expected upper limits on the signal cross-section.

Finally, in section 7.7.3.1, it has been observed that the MC background samples do not model well the number of data events in the blinded CRs of the ABCD plane. The data in these regions amount to only about 40% of the MC background contributions. It is thus expected, in the unblinded data, to have the same scale factor between data and MC in the ABCD plane. The expected upper limits on the signal cross-section computed in the data will consequently be lower, providing a better signal model rejection.

7.9.3.3 Effect of the systematic uncertainties on the signal cross-section upper limits

This last subsection presents the effect of the systematic uncertainties of both the SR yield and CRs yields on the signal cross-section upper limits. First, to measure the effect of the SR yield systematic uncertainty σ^{SR} , upper limits on the signal cross-sections are re-computed with the statistical framework without the constraint term on the systematic uncertainty (i.e. $\sigma^{\text{SR}} = 0$) and compared to the nominal upper limits. Tab. 7.16 presents the obtained relative variations of the limits in percent between the two computations for each signal and each limit. The variations are observed to be of the order of the percent, always below 5 %, for the nominal expected limits. The differences are greater for the expected band, enlarging it, the systematic uncertainty on the SR yield thus increases the standard deviation of $\hat{\mu}$. Moreover, as expected, the expected limits are consistently lower without the systematic uncertainties, this effect can be seen as a loss of information on the signal yield when considering signal systematics.

Table 7.16: Relative variations in percent on the signal cross-section upper limits with or without systematics on the signal strength.

Signal model	Variation in the expected limits (%)				
	$\sigma_{\text{up,exp}}^{-2\sigma}$	$\sigma_{\text{up,exp}}^{-1\sigma}$	$\sigma_{\text{up,exp}}$	$\sigma_{\text{up,exp}}^{+1\sigma}$	$\sigma_{\text{up,exp}}^{+2\sigma}$
Ld10_rho20_pi5_Zp1500_15	0.95	1.55	3.37	7.05	12.0
Ld10_rho20_pi5_Zp1500_150	1.38	2.23	4.87	9.63	16.4
Ld10_rho20_pi5_Zp3000_15	0.43	0.53	1.19	2.47	4.11
Ld10_rho20_pi5_Zp3000_150	0.35	0.52	1.06	2.18	3.58
Ld20_rho40_pi10_Zp1500_15	0.96	1.61	3.39	6.99	12.0
Ld20_rho40_pi10_Zp1500_150	0.88	1.41	2.89	5.95	10.1
Ld20_rho40_pi10_Zp3000_15	0.18	0.39	1.00	2.08	3.50
Ld20_rho40_pi10_Zp3000_150	0.23	0.33	0.70	1.47	2.32
Ld40_rho80_pi20_Zp1500_15	0.95	1.39	3.14	6.47	11.1
Ld40_rho80_pi20_Zp1500_150	0.70	1.17	2.41	4.89	8.44
Ld40_rho80_pi20_Zp3000_15	0.37	0.54	1.13	2.42	4.10
Ld40_rho80_pi20_Zp3000_150	0.12	0.27	0.51	1.11	1.70

Second, to understand the effect of the systematic uncertainties on the signal yields in the CRs, and thus on the signal leakages, a similar procedure is applied. The upper limits on the signal cross-sections are re-computed by manually changing the signal leakage, supplied to the statistical model, to the maximal

upward or downward variations of these quantities. The upward and downward variations were given in the previous section (Tab. 7.12) and computed by propagating the systematic uncertainties on the signal yields. The varied signal leakages are given by:

$$q_X \rightarrow q_X + q_X^{\text{up}} \text{ or } q_X - q_X^{\text{down}} \quad (7.50)$$

The obtained varied upper limits are then compared to the limits obtained with the nominal signal leakages. Tab. 7.17 shows the obtained relative variations in percent on the signal cross-section upper limits between the nominal computation and the one with all signal leakages varying upward. The variations are observed to be very small, under 0.5 %, and generally one order of magnitude below the variations obtained in Tab. 7.16. Similar results are obtained with the signal leakages varying downward. The variations on the signal leakages are thus considered to be negligible. This motivates the fact that these systematic uncertainties are not included in the statistical framework.

Table 7.17: Relative variations in percent on the signal cross-section upper limits obtained with the maximal upward variations of the signal leakages q_X .

Signal model	Variation in the expected limits (%)				
	$\sigma_{\text{up,exp}}^{-2\sigma}$	$\sigma_{\text{up,exp}}^{-1\sigma}$	$\sigma_{\text{up,exp}}$	$\sigma_{\text{up,exp}}^{+1\sigma}$	$\sigma_{\text{up,exp}}^{+2\sigma}$
Ld10_rho20_pi5_Zp1500_15	0.31	0.23	0.32	0.28	0.29
Ld10_rho20_pi5_Zp1500_150	0.30	0.28	0.30	0.25	0.27
Ld10_rho20_pi5_Zp3000_15	0.14	0.11	0.07	0.10	0.14
Ld10_rho20_pi5_Zp3000_150	0.07	0.05	0.07	0.08	0.07
Ld20_rho40_pi10_Zp1500_15	0.41	0.30	0.41	0.37	0.34
Ld20_rho40_pi10_Zp1500_150	0.16	0.12	0.16	0.16	0.13
Ld20_rho40_pi10_Zp3000_15	0.35	0.26	0.27	0.31	0.30
Ld20_rho40_pi10_Zp3000_150	0.00	0.08	0.06	0.04	0.06
Ld40_rho80_pi20_Zp1500_15	0.27	0.29	0.34	0.27	0.28
Ld40_rho80_pi20_Zp1500_150	0.18	0.06	0.13	0.12	0.12
Ld40_rho80_pi20_Zp3000_15	0.18	0.27	0.19	0.19	0.22
Ld40_rho80_pi20_Zp3000_150	0.12	0.00	0.06	0.04	0.06

7.9.3.4 Results in the context of the global search for a new vector mediator Z'

In this analysis, the Z' mediator is assumed to predominantly decay to dark quarks but, as it needs to be produced in SM pp collisions, it necessarily also decay to SM quarks. The Z' has thus two decay channels represented by the two Feynman diagrams shown in Fig. 7.49. To avoid generating useless signal events, to which the analysis would not be sensible, the Z' is set to decay 99.9 % of the times to dark quarks in

the MC signal samples but in fact there are no strong theory priors to have such an extreme case. The branching ratios of the Z' , and thus its couplings g_q and g_D , can be varied freely.

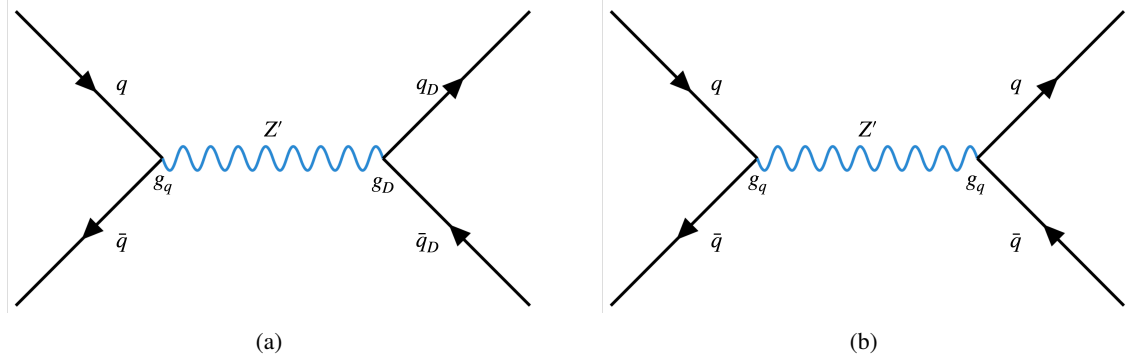


Figure 7.49: Feynman diagram for the pair production of dark quarks (a) or SM quarks (b) through a Z' vector mediator.

Several previous analyses searched for the production of a new vector mediator Z' decaying hadronically [135–137] by analysing the dijet invariant mass spectrum and looking for a bump that would provide evidences of a resonant decay of a new massive particle. Fig. 7.50 presents a summary plot of the exclusion limits on g_q as a function of $m_{Z'}$ obtained by several ATLAS analyses searching for a Z' mediator in different decay channels and where the Z' is assumed to only decay to SM quarks. The production process of EJs through a Z' is consequently also constrained by such analysis because of the decay of the Z' to SM quarks. However, for small values of coupling g_q , lower than approximately 0.03, there are no constraints extracted by these analyses. In such a region of phase-space, the EJ signal models are not excluded by the $Z' \rightarrow q\bar{q}$ searches.

In a very preliminary study and for small values of g_q , the EJ analysis can be shown to be able to put strong constraint on the EJ production cross-section and therefore on the couplings g_q , related to the global Z' production cross-section, and g_D , linked to the decay rate of the Z' to dark quarks. Using MADGRAPH [146], the production of a Z' decaying to dark quarks can be generated by setting directly the couplings of the Z' , g_q and g_D ¹¹. The generator then provides the production cross-section of the Z' to dark quarks. Thus, for the model parameters corresponding to the signal samples and for several values of g_q and g_D , it is possible to compare the obtained theoretical cross-sections to the measured expected limits. The other parameters impacting the cross-section are: $m_{Z'}$ and m_{q_D} . The mass of the dark quark being at the GeV scale, it has a very small impact due to the large mass hierarchy with the Z' , such that only one value of m_{q_D} is used. As an example, with $m_{Z'} = 1.5$ TeV and with $g_q = 0.03$ and $g_D = 0.1$, the theoretical cross-section for $pp \rightarrow Z' \rightarrow q_D \bar{q}_D$ equals 17.5 fb, one order of magnitude above the expected limits for the $m_{Z'} = 1.5$ TeV models obtained in this thesis. With $m_{Z'} = 1.5$ TeV and $g_q = 0.03$, the EJ analysis is able to exclude values of g_D down to at least 0.1 based on this preliminary study alone. For models with $m_{Z'} = 3$ TeV, the analysis is not able to exclude any values of g_D with $g_q = 0.03$. However, for such mass, the constraints from the dijet searches are weaker, no exclusion is obtained for g_q up to around 0.1. For such a value of g_q and with $m_{Z'} = 3$ TeV, the EJ analysis is then able to exclude values of g_D down to around 0.2 (with $g_D = 0.2$, the cross-section equals 2.2 fb). These results were obtained in a preliminary study with a coarse scan of the couplings, g_q and g_D , values. With the final results that will be obtained in the unblinded data, a finer scan could be made to obtain an exclusion contour in the (g_q, g_D) plane.

¹¹ Compared to PYTHIA where only the Z' decay width and the branching ratios to SM quarks and dark quarks can be set, which is equivalent to setting the couplings.

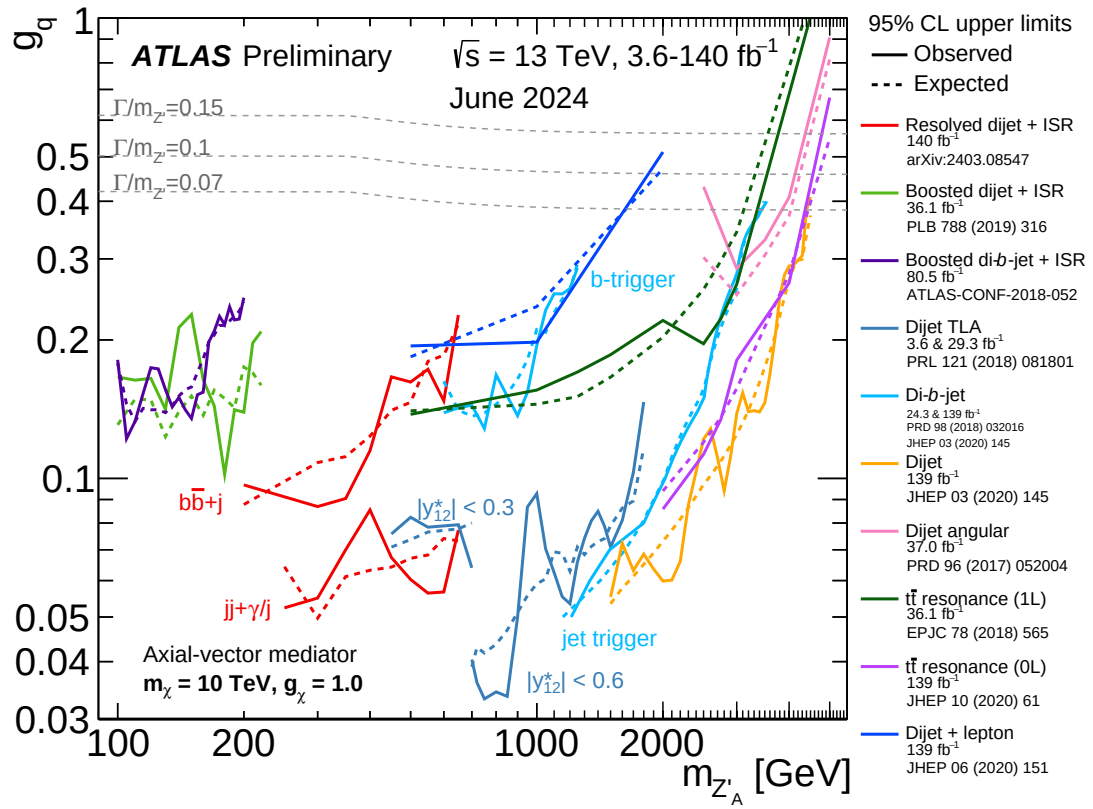


Figure 7.50: Summary plot of the 95% CL upper limits on g_q as a function of the vector mediator mass $m_{Z'}$ from several analyses varying the targetted decay channel of the Z' [145].

7.10 Conclusion

This Chapter presented an exotic search for emerging jets produced by a vector mediator in the s-channel. It is the first time that this topology of EJs is investigated by either ATLAS or CMS. Previous analyses searching for EJs were conducted but only in the bi-fundamental t-channel topology by CMS [36, 37]. The analysis strategy revolved around the discrimination of the EJs against the SM jets to define a SR enriched in signal events with a high background rejection. Using a data-driven background estimation method validated in blinded data, the contribution of the dijet+ $t\bar{t}$ background in the SR is consistently estimated. The background estimation is then used to test for the presence of an excess of events in the SR. The statistical interpretation was put in place and predicted results on the expected signal cross-section limits were computed from MC background simulations. This analysis strategy is ready for approval within the ATLAS Collaboration, which will allow access to the unblinded data for the final statistical interpretation. The analysis should then lead to the publication of a paper. I strongly contributed to the development of the analysis software in terms of reconstructions, calibrations, selections and observable computations. Furthermore, I solely established the exposed analysis strategy based on the large- R jet trigger from the signal discrimination studies to the statistical interpretation.

This analysis is the first throw at studying EJ models in a vector mediator s-channel topology within ATLAS or CMS. It is, however, part of a recent and broad effort to search for dark QCD within the community with several published results investigating the different signatures of dark jets (prompt dark jets [33], semi-visible jets [34, 35], EJs with a bi-fundamental mediator [36, 37]). As this analysis is a first iteration for a search for EJs with a vector mediator in ATLAS, further improvements are imaginable for the full Run 3 dataset and for Run 4, in addition to the increase of luminosity. In particular the upgrade of ATLAS' inner detector planned for Run 4 and the HL-LHC will provide a greater η coverage and better track and vertex reconstruction efficiencies. Better jet reconstruction can also be beneficial, investigating larger opening angle and dedicated jet constituents could provide important improvements for reconstructing EJs. Strategies based on machine learning techniques to identify EJs are also currently investigated for this analysis. They show very good performances and continuous development of these techniques are thus expected to provide better background rejection and signal acceptance for the current and future iterations of the analysis.

CHAPTER 8

GENERAL CONCLUSION

The SM of particle physics is a very successful theory that allowed to predict many particles and their interactions that are well corroborated by several particle physics experiments. However, it is acknowledged that the SM is missing important pieces and in particular a candidate particle to explain the astrophysical and cosmological evidences for DM. This thesis explored a dark sector theory that can predict particle candidates for dark matter. The dark sectors have only recently started to be broadly searched for by the LHC experiments and several results are already available. The corresponding analyses are often the first of their kind and require a precise knowledge of the detectors' performances as the dark sectors usually yield exotic signatures involving uncommon signatures like displaced objects. The next iterations of these analyses are thus expected to be able to improve their sensitivities by exploiting upgrades in event reconstruction, detector technologies and machine learning techniques. Moreover, due to the large range of potential signatures produced by the dark sectors, there are still large portions of phase space to cover. The dark sectors are a very dynamic field of study which holds potential discoveries in the future exploitation of the LHC and the HL-LHC.

In my thesis, I participated in the development of a new machine learning based technique to correct the energy and mass of large- R UFO jets. Using a DNN, the large- R jets energy and mass are simultaneously calibrated by predicting their average responses in the detector. My work consisted in the development of the DNN architecture and training to obtain a calibration that outperforms the standard calibration currently used in ATLAS. The DNN calibration was validated in a series of tests that allowed to ensure the calibration to be physically consistent and able to generalise to different jet topologies. Ultimately, this work led to the publication of a paper in the journal MLST and this new calibration method for large- R UFO jets will replace the current standard calibration in ATLAS such that it can be used in future analyses.

In parallel, I also participated in the design of a new analysis in ATLAS aiming at EJs, a particular setup of dark QCD that predicts unstable dark particles with long lifetimes. This analysis is the first that investigates EJs in an s-channel dijet production at the LHC. I contributed in a large range of technical tasks that were crucial for the progression of the analysis and I solely developed one of the analysis strategies from the signal discrimination studies to define a signal region, through the background estimation method and up to the statistical interpretation. However, due to a lack of time, this analysis strategy is not yet approved by the ATLAS Collaboration such that only the expected exclusion limits on the signal cross-section computed from the MC background simulations are provided. The procedure of approval should however start shortly giving access to the unblinded data for the final statistical interpretation and should ultimately lead to the publication of a paper.

BIBLIOGRAPHY

- [1] M. K. Gaillard, P. D. Grannis and F. J. Sciulli, *The standard model of particle physics*, *Reviews of Modern Physics* **71** (1999) S96, ISSN: 1539-0756, URL: <http://dx.doi.org/10.1103/RevModPhys.71.S96> (cit. on p. 9).
- [2] J. Drees, *Review of final LEP results or a tribute to LEP*, *International Journal of Modern Physics A* **17** (2002) 3259, ISSN: 1793-656X, URL: <http://dx.doi.org/10.1142/S0217751X02012727> (cit. on p. 9).
- [3] D. Bandurin et al., *Review of physics results from the Tevatron*, *International Journal of Modern Physics A* **30** (2015) 1541001, ISSN: 1793-656X, URL: <http://dx.doi.org/10.1142/S0217751X15410018> (cit. on p. 9).
- [4] M. Klein and R. Yoshida, *Collider physics at HERA*, *Progress in Particle and Nuclear Physics* **61** (2008) 343, ISSN: 0146-6410, URL: <https://www.sciencedirect.com/science/article/pii/S0146641008000392> (cit. on p. 9).
- [5] O. Gonzalez, *LHC Results Highlights (CLASHEP 2013)*, 2013, arXiv: [1307.6992 \[hep-ex\]](#) (cit. on p. 9).
- [6] D. Galbraith and C. Burgard, *Diagram of the standard model of particle physics*, (2012), URL: <https://texample.net/tikz/examples/model-physics/> (cit. on p. 9).
- [7] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (cit. on p. 10).
- [8] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, arXiv: [1207.7214 \[hep-ex\]](#) (cit. on p. 10).
- [9] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30, arXiv: [1207.7235 \[hep-ex\]](#) (cit. on p. 10).
- [10] F. Gross et al., *50 Years of quantum chromodynamics*, *The European Physical Journal C* **83** (2023), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-023-11949-2> (cit. on p. 11).
- [11] *Feynman diagrams of QCD interactions*, WIKIPEDIA (2017), URL: https://fr.m.wikipedia.org/wiki/Diagramme_de_Feynman (cit. on p. 11).
- [12] P. D. Group, P. A. Zyla and R. M. Barnett, *Review of Particle Physics*, *Progress of Theoretical and Experimental Physics* **2020** (2020) 083C01, ISSN: 2050-3911, eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>, URL: <https://doi.org/10.1093/ptep/ptaa104> (cit. on p. 12).

- [13] A. Collaboration, *A precise determination of the strong-coupling constant from the recoil of Z bosons with the ATLAS experiment at $\sqrt{s} = 8$ TeV*, 2023, arXiv: [2309.12986 \[hep-ex\]](#) (cit. on p. 12).
- [14] K. G. Wilson, *Confinement of Quarks*, *Phys. Rev. D* **10** (1974) 2445, ed. by J. C. Taylor (cit. on p. 12).
- [15] S. Höche, *Introduction to parton-shower event generators*, 2015, arXiv: [1411.4085 \[hep-ph\]](#) (cit. on p. 14).
- [16] G. Bertone and D. Hooper, *History of dark matter*, *Reviews of Modern Physics* **90** (2018), ISSN: 1539-0756, URL: <http://dx.doi.org/10.1103/RevModPhys.90.045002> (cit. on p. 16).
- [17] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*, *Astrophysical Journal* **86** (1937) 217 (cit. on p. 16).
- [18] V. C. Rubin and J. Ford W. Kent, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, *Astrophysical Journal* **159** (1970) 379 (cit. on p. 16).
- [19] M. S. Roberts and A. H. Rots, *Comparison of Rotation Curves of Different Galaxy Types*, *Astronomy and Astrophysics* **26** (1973) 483 (cit. on p. 16).
- [20] K. G. Begeman, A. H. Broeils and R. H. Sanders, *Extended rotation curves of spiral galaxies: dark haloes and modified dynamics*, *Monthly Notices of the Royal Astronomical Society* **249** (1991) 523, ISSN: 0035-8711, eprint: <https://academic.oup.com/mnras/article-pdf/249/3/523/18160929/mnras249-0523.pdf>, URL: <https://doi.org/10.1093/mnras/249.3.523> (cit. on p. 17).
- [21] D. Clowe, A. Gonzalez and M. Markevitch, *Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter*, *The Astrophysical Journal* **604** (2004) 596, ISSN: 1538-4357, URL: <http://dx.doi.org/10.1086/381970> (cit. on p. 16).
- [22] M. Bradac et al., *Revealing the Properties of Dark Matter in the Merging Cluster MACS J0025.4-1222*, *The Astrophysical Journal* **687** (2008) 959, ISSN: 1538-4357, URL: <http://dx.doi.org/10.1086/591246> (cit. on p. 17).
- [23] G. R. Blumenthal, S. M. Faber, J. R. Primack and M. J. Rees, *Formation of galaxies and large-scale structure with cold dark matter.*, *Nature* **311** (1984) 517 (cit. on p. 17).
- [24] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astronomy & Astrophysics* **641** (2020) A6, ISSN: 1432-0746, URL: <http://dx.doi.org/10.1051/0004-6361/201833910> (cit. on p. 17).
- [25] P. Fayet and S. Ferrara, *Supersymmetry*, *Physics Reports* **32** (1977) 249, ISSN: 0370-1573, URL: <https://www.sciencedirect.com/science/article/pii/0370157377900667> (cit. on p. 18).

- [26] A. Canepa, *Searches for supersymmetry at the Large Hadron Collider*, *Reviews in Physics* **4** (2019) 100033, ISSN: 2405-4283, URL: <https://www.sciencedirect.com/science/article/pii/S2405428318300091> (cit. on p. 18).
- [27] L. Roszkowski, E. M. Sessolo and S. Trojanowski, *WIMP dark matter candidates and searches—current status and future prospects*, *Reports on Progress in Physics* **81** (2018) 066201, ISSN: 1361-6633, URL: <http://dx.doi.org/10.1088/1361-6633/aab913> (cit. on p. 18).
- [28] C. B. Adams et al., *Axion Dark Matter*, 2023, arXiv: [2203.14923 \[hep-ex\]](#) (cit. on p. 18).
- [29] ATLAS Collaboration, *Searching for Dark Matter with the ATLAS detector*, (2019), URL: <https://atlas.cern/updates/feature/dark-matter> (cit. on p. 20).
- [30] CMS Collaboration, *Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **97** (2018) 092005, arXiv: [1712.02345 \[hep-ex\]](#) (cit. on p. 20).
- [31] K. Petraki and R. R. Volkas, *Review of asymmetric dark matter*, *International Journal of Modern Physics A* **28** (2013) 1330028, ISSN: 1793-656X, URL: <http://dx.doi.org/10.1142/S0217751X13300287> (cit. on p. 21).
- [32] G. Albouy et al., *Theory, phenomenology, and experimental avenues for dark showers: a Snowmass 2021 report*, *The European Physical Journal C* **82** (2022), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-022-11048-8> (cit. on p. 21).
- [33] ATLAS Collaboration, *Search for resonant production of dark quarks in the dijet final state with the ATLAS detector*, *JHEP* **02** (2023) 128, arXiv: [2311.03944 \[hep-ex\]](#) (cit. on pp. 22, 190).
- [34] CMS Collaboration, *Search for resonant production of strongly coupled dark matter in proton–proton collisions at 13 TeV*, *JHEP* **06** (2022) 156, arXiv: [2112.11125 \[hep-ex\]](#) (cit. on pp. 22, 190).
- [35] ATLAS Collaboration, *Search for non-resonant production of semi-visible jets using Run 2 data in ATLAS*, *Phys. Lett. B* **848** (2024) 138324, arXiv: [2305.18037 \[hep-ex\]](#) (cit. on pp. 22, 190).
- [36] CMS Collaboration, *Search for new particles decaying to a jet and an emerging jet*, *JHEP* **02** (2019) 179, arXiv: [1810.10069 \[hep-ex\]](#) (cit. on pp. 22, 190).
- [37] CMS Collaboration, *Search for new physics with emerging jets in proton–proton collisions at $\sqrt{s} = 13$ TeV*, (2024), arXiv: [2403.01556 \[hep-ex\]](#) (cit. on pp. 22, 190).
- [38] M. Park and M. Zhang, *Tagging a jet from a dark sector with jet substructures at colliders*, *Physical Review D* **100** (2019), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.100.115009> (cit. on p. 23).
- [39] P. Schwaller, D. Stolarski and A. Weiler, *Emerging jets*, *Journal of High Energy Physics* **2015** (2015), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP05\(2015\)059](http://dx.doi.org/10.1007/JHEP05(2015)059) (cit. on pp. 24, 25).
- [40] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003 (cit. on p. 27).

- [41] CMS Collaboration, *The CMS Experiment at the CERN LHC*, [JINST 3 \(2008\) S08004](#) (cit. on p. 27).
- [42] K. Aamodt et al., *The ALICE experiment at the CERN LHC*, [JINST 3 \(2008\) S08002](#) (cit. on p. 27).
- [43] A. A. Alves Jr. et al., *The LHCb Detector at the LHC*, [JINST 3 \(2008\) S08005](#) (cit. on p. 27).
- [44] I. Tsukerman for the ATLAS and CMS Collaborations, *Measurements of the Higgs boson by ATLAS and CMS*, [Journal of Physics: Conference Series 1390 \(2019\) 012030](#), URL: <https://dx.doi.org/10.1088/1742-6596/1390/1/012030> (cit. on p. 27).
- [45] E. Lopienska, *The CERN accelerator complex, layout in 2022*, (2022), General Photo, URL: <https://cds.cern.ch/record/2800984> (cit. on p. 28).
- [46] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider: A Description of the Detector Configuration for Run 3*, (2023), arXiv: [2305.16623 \[physics.ins-det\]](#) (cit. on pp. 30, 33–37).
- [47] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 1*, ATLAS-TDR-4; CERN-LHCC-97-016, 1997, URL: <https://cds.cern.ch/record/331063> (cit. on p. 30).
- [48] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 2*, ATLAS-TDR-5, CERN-LHCC-97-017, 1997, URL: <https://cds.cern.ch/record/331064> (cit. on p. 30).
- [49] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, ATLAS-TDR-2; CERN-LHCC-96-041, 1996, URL: <https://cds.cern.ch/record/331061> (cit. on p. 30).
- [50] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, ATLAS-TDR-3; CERN-LHCC-96-042, 1996, URL: <https://cds.cern.ch/record/331062> (cit. on p. 30).
- [51] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, ATLAS-TDR-10; CERN-LHCC-97-022, CERN, 1997, URL: <https://cds.cern.ch/record/331068> (cit. on p. 30).
- [52] ATLAS Collaboration, *ATLAS Magnet System: Magnet Project Technical Design Report, Volume 1*, ATLAS-TDR-6; CERN-LHCC-97-018, 1997, URL: <https://cds.cern.ch/record/338080> (cit. on p. 30).
- [53] ATLAS Collaboration, *ATLAS Barrel Toroid: Magnet Project Technical Design Report, Volume 2*, ATLAS-TDR-7; CERN-LHCC-97-019, 1997, URL: <https://cds.cern.ch/record/331065> (cit. on p. 30).
- [54] ATLAS Collaboration, *ATLAS End-Cap Toroids: Magnet Project Technical Design Report, Volume 3*, ATLAS-TDR-8; CERN-LHCC-97-020, CERN, 1997, URL: <https://cds.cern.ch/record/331066> (cit. on p. 30).

- [55] ATLAS Collaboration, *ATLAS Central Solenoid: Magnet Project Technical Design Report, Volume 4*, ATLAS-TDR-9; CERN-LHCC-97-021, CERN, 1997, URL: <https://cds.cern.ch/record/331067> (cit. on p. 30).
- [56] *CMS coordinate system*, (2022), URL: https://tikz.net/axis3d_cms/ (cit. on p. 31).
- [57] W. Herr and B. Muratori, *Concept of luminosity*, (2006), URL: <https://cds.cern.ch/record/941318> (cit. on p. 38).
- [58] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-3 of the LHC*, (2023), URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3> (cit. on pp. 39–41, 116).
- [59] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, *JINST* **15** (2020) P04003, arXiv: 1911.04632 [physics.ins-det] (cit. on pp. 39, 113).
- [60] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159, arXiv: 1410.3012 [hep-ph] (cit. on pp. 43, 79, 113).
- [61] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, arXiv: 1905.09127 [hep-ph] (cit. on pp. 43, 79, 114).
- [62] J. Bellm et al., *Herwig 7.1 Release Note*, (2017), arXiv: 1705.06919 [hep-ph] (cit. on p. 43).
- [63] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 43).
- [64] S. Agostinelli et al., *Geant4—a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250, ISSN: 0168-9002 (cit. on p. 43).
- [65] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 43).
- [66] ATLAS Collaboration, *Athena*, <https://doi.org/10.5281/zenodo.4772550> (2021) (cit. on p. 44).
- [67] ATLAS Collaboration, *ATLAS tracking parameters*, (2023), URL: <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idooverview/> (cit. on p. 47).
- [68] ATLAS Collaboration, *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*, (2023), arXiv: 2308.09471 [hep-ex] (cit. on pp. 48, 49).
- [69] ATLAS Collaboration, *Performance of the reconstruction of large impact parameter tracks in the inner detector of ATLAS*, *Eur. Phys. J. C* **83** (2023) 1081, arXiv: 2304.12867 [hep-ex] (cit. on pp. 48–50).
- [70] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987) 444, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874> (cit. on p. 49).

- [71] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, ATL-PHYS-PUB-2019-015, 2019, URL: <https://cds.cern.ch/record/2670380> (cit. on p. 51).
- [72] ATLAS Collaboration, *Performance of vertex reconstruction algorithms for detection of new long-lived particle decays within the ATLAS inner detector*, ATL-PHYS-PUB-2019-013, 2019, URL: <https://cds.cern.ch/record/2669425> (cit. on pp. 52–54).
- [73] V. Kostyukhin, *VKalVrt - package for vertex reconstruction in ATLAS.*, tech. rep., revised version number 1 submitted on 2003-09-24 11:10:53: CERN, 2003, URL: <https://cds.cern.ch/record/685551> (cit. on p. 52).
- [74] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896, arXiv: [1111.6097 \[hep-ph\]](https://arxiv.org/abs/1111.6097) (cit. on p. 55).
- [75] M. Cacciari, G. P. Salam and G. Soyez, *The anti-kt jet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008) 063, ISSN: 1029-8479, URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063> (cit. on p. 55).
- [76] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160, arXiv: [hep-ph/9305266](https://arxiv.org/abs/hep-ph/9305266) (cit. on p. 56).
- [77] Y. Dokshitzer, G. Leder, S. Moretti and B. Webber, *Better jet clustering algorithms*, *Journal of High Energy Physics* **1997** (1997) 001, ISSN: 1029-8479, URL: <http://dx.doi.org/10.1088/1126-6708/1997/08/001> (cit. on p. 56).
- [78] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) 490, arXiv: [1603.02934 \[hep-ex\]](https://arxiv.org/abs/1603.02934) (cit. on pp. 57, 58).
- [79] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466, arXiv: [1703.10485 \[hep-ex\]](https://arxiv.org/abs/1703.10485) (cit. on p. 60).
- [80] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 689, arXiv: [2007.02645 \[hep-ex\]](https://arxiv.org/abs/2007.02645) (cit. on pp. 60, 69, 70, 77, 80, 85, 94, 169).
- [81] ATLAS Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, ATL-PHYS-PUB-2017-015, 2017, URL: <https://cds.cern.ch/record/2275636> (cit. on pp. 62, 63).
- [82] ATLAS Collaboration, *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions*, *Eur. Phys. J. C* **81** (2021) 334, arXiv: [2009.04986 \[hep-ex\]](https://arxiv.org/abs/2009.04986) (cit. on pp. 65, 66, 69).
- [83] P. Berta, M. Spousta, D. W. Miller and R. Leitner, *Particle-level pileup subtraction for jets and jet shapes*, *Journal of High Energy Physics* **2014** (2014), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP06\(2014\)092](http://dx.doi.org/10.1007/JHEP06(2014)092) (cit. on p. 67).
- [84] M. Cacciari, G. P. Salam and G. Soyez, *SoftKiller, a particle-level pileup removal method*, *The European Physical Journal C* **75** (2015), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-015-3267-2> (cit. on p. 68).

- [85] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, *Journal of High Energy Physics* **2010** (2010), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP02\(2010\)084](http://dx.doi.org/10.1007/JHEP02(2010)084) (cit. on p. 68).
- [86] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, arXiv: [1402.2657 \[hep-ph\]](#) (cit. on p. 69).
- [87] ATLAS Collaboration, *New techniques for jet calibration with the ATLAS detector*, *Eur. Phys. J. C* **83** (2023) 761, arXiv: [2303.17312 \[hep-ex\]](#) (cit. on pp. 69, 71, 72, 96).
- [88] A. Cukierman and B. Nachman, *Mathematical Properties of Numerical Inversion for Jet Calibrations*, *Nucl. Instrum. Meth. A* **858** (2017) 1, arXiv: [1609.05195 \[physics.data-an\]](#) (cit. on pp. 72, 77).
- [89] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 135, arXiv: [1807.09477 \[hep-ex\]](#) (cit. on p. 73).
- [90] ATLAS Collaboration, *Simultaneous energy and mass calibration of large-radius jets with the ATLAS detector using a deep neural network*, *Machine Learning: Science and Technology* **5** (2024) 035051, URL: <https://dx.doi.org/10.1088/2632-2153/ad611e> (cit. on pp. 74, 110).
- [91] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J. C* **73** (2013) 2304, arXiv: [1112.6426 \[hep-ex\]](#) (cit. on p. 77).
- [92] ATLAS Collaboration, *Generalized Numerical Inversion: A Neutral Network Approach to Jet Calibration*, ATL-PHYS-PUB-2018-013, 2018, URL: <https://cds.cern.ch/record/2630972> (cit. on p. 78).
- [93] NNPDF Collaboration, R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: [1207.1303 \[hep-ph\]](#) (cit. on pp. 79, 113, 116).
- [94] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419> (cit. on pp. 79, 113).
- [95] L. Randall and R. Sundrum, *Large Mass Hierarchy from a Small Extra Dimension*, *Physical Review Letters* **83** (1999) 3370, arXiv: [9905221 \[hep-ph\]](#) (cit. on p. 79).
- [96] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, *JHEP* **03** (2008) 038, arXiv: [0709.1027 \[hep-ph\]](#) (cit. on pp. 79, 114).
- [97] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev. D* **93** (2016) 033006, arXiv: [1506.07443 \[hep-ph\]](#) (cit. on p. 79).
- [98] J.-C. Winter, F. Krauss and G. Soff, *A modified cluster-hadronisation model*, *Eur. Phys. J. C* **36** (2004) 381, arXiv: [hep-ph/0311085](#) (cit. on p. 79).
- [99] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, arXiv: [0710.3820 \[hep-ph\]](#) (cit. on pp. 79, 116).

- [100] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: <https://cds.cern.ch/record/2206965> (cit. on pp. 79, 116).
- [101] H. W. Kurt Hornik Maxwell Stinchcombe, *Multilayer feedforward networks are universal approximators*, *Neural Networks* **2** (1989) 359, ISSN: 0893-6080, URL: <https://www.sciencedirect.com/science/article/pii/0893608089900208> (cit. on p. 81).
- [102] M. Abadi et al., ‘TensorFlow: A System for Large-Scale Machine Learning’, *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, USENIX Association, 2016 265, ISBN: 9781931971331 (cit. on pp. 82, 90).
- [103] S. Cheong, A. Cukierman, B. Nachman, M. Safdari and A. Schwartzman, *Parametrizing the detector response with neural networks*, *JINST* **15** (2020) P01030, arXiv: [1910.03773](https://arxiv.org/abs/1910.03773) [[physics.data-an](#)] (cit. on p. 84).
- [104] ATLAS Collaboration, *Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC*, *Eur. Phys. J. C* **79** (2019) 375, arXiv: [1808.07858](https://arxiv.org/abs/1808.07858) [[hep-ex](#)] (cit. on p. 85).
- [105] J. Thaler and L.-T. Wang, *Strategies to identify boosted tops*, *JHEP* **07** (2008) 092, arXiv: [0806.0023](https://arxiv.org/abs/0806.0023) [[hep-ph](#)] (cit. on pp. 86, 139, 145).
- [106] A. J. Larkoski, G. P. Salam and J. Thaler, *Energy correlation functions for jet substructure*, *JHEP* **06** (2013) 108, arXiv: [1305.0007](https://arxiv.org/abs/1305.0007) [[hep-ph](#)] (cit. on pp. 86, 139, 145).
- [107] A. J. Larkoski, I. Moult and D. Neill, *Power counting to better jet observables*, *JHEP* **12** (2014) 009, arXiv: [1409.6298](https://arxiv.org/abs/1409.6298) [[hep-ph](#)] (cit. on pp. 86, 139).
- [108] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015, arXiv: [1011.2268](https://arxiv.org/abs/1011.2268) [[hep-ph](#)] (cit. on pp. 86, 139, 145).
- [109] A. Larkoski, D. Neill and J. Thaler, *Jet shapes with the broadening axis*, *JHEP* **04** (2014) 17, arXiv: [1401.2158](https://arxiv.org/abs/1401.2158) [[hep-ph](#)] (cit. on pp. 86, 139).
- [110] K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition*, *CoRR abs/1512.03385* (2015), arXiv: [1512.03385](https://arxiv.org/abs/1512.03385) (cit. on p. 88).
- [111] D. Misra, *Mish: A Self Regularized Non-Monotonic Neural Activation Function*, *CoRR* (2019), arXiv: [1908.08681](https://arxiv.org/abs/1908.08681) (cit. on p. 89).
- [112] L. Liu et al., *On the Variance of the Adaptive Learning Rate and Beyond*, *CoRR abs/1908.03265* (2019), arXiv: [1908.03265](https://arxiv.org/abs/1908.03265) (cit. on p. 90).
- [113] S. R. Dubey et al., *diffGrad: An Optimization Method for Convolutional Neural Networks*, *CoRR abs/1909.11015* (2019), arXiv: [1909.11015](https://arxiv.org/abs/1909.11015) (cit. on p. 90).
- [114] J. Pivarski et al., *scikit-hep/uproot: 3.12.0*, version 3.12.0, 2020, URL: <https://doi.org/10.5281/zenodo.3952728> (cit. on p. 90).
- [115] P.-A. Delsart and G. Albouy, *DNN jet calibration development software*, version 1, URL: <https://gitlab.cern.ch/galbouy/jetcalibdmn> (cit. on p. 90).

- [116] A. Savitzky and M. J. E. Golay, *Smoothing and Differentiation of Data by Simplified Least Squares Procedures.*, *Analytical Chemistry* **36** (1964) 1627, ISSN: 0003-2700, URL: <https://doi.org/10.1021/ac60214a047> (cit. on p. 93).
- [117] N. Nanda, L. Chan, T. Lieberum, J. Smith and J. Steinhardt, ‘Progress measures for grokking via mechanistic interpretability’, *The Eleventh International Conference on Learning Representations*, 2023, arXiv: [2301.05217](https://arxiv.org/abs/2301.05217), URL: <https://openreview.net/forum?id=9XFSbDPmdW> (cit. on p. 106).
- [118] S. M. Lundberg and S.-I. Lee, ‘A Unified Approach to Interpreting Model Predictions’, *Advances in Neural Information Processing Systems*, vol. 30, 2017, arXiv: [1705.07874](https://arxiv.org/abs/1705.07874) (cit. on p. 109).
- [119] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, arXiv: [0808.3674](https://arxiv.org/abs/0808.3674) [[hep-ph](#)] (cit. on p. 114).
- [120] F. Buccioni et al., *OpenLoops 2*, *Eur. Phys. J. C* **79** (2019) 866, arXiv: [1907.13071](https://arxiv.org/abs/1907.13071) [[hep-ph](#)] (cit. on p. 114).
- [121] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, arXiv: [1111.5206](https://arxiv.org/abs/1111.5206) [[hep-ph](#)] (cit. on p. 114).
- [122] A. Denner, S. Dittmaier and L. Hofer, *COLLIER: A fortran-based complex one-loop library in extended regularizations*, *Comput. Phys. Commun.* **212** (2017) 220, arXiv: [1604.06792](https://arxiv.org/abs/1604.06792) [[hep-ph](#)] (cit. on p. 114).
- [123] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, arXiv: [1111.1220](https://arxiv.org/abs/1111.1220) [[hep-ph](#)] (cit. on p. 114).
- [124] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, *JHEP* **04** (2013) 027, arXiv: [1207.5030](https://arxiv.org/abs/1207.5030) [[hep-ph](#)] (cit. on p. 114).
- [125] S. Catani, F. Krauss, B. R. Webber and R. Kuhn, *QCD Matrix Elements + Parton Showers*, *JHEP* **11** (2001) 063, arXiv: [hep-ph/0109231](https://arxiv.org/abs/hep-ph/0109231) (cit. on p. 114).
- [126] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, arXiv: [0903.1219](https://arxiv.org/abs/0903.1219) [[hep-ph](#)] (cit. on p. 114).
- [127] NNPDF Collaboration, R. D. Ball et al., *Parton distributions for the LHC run II*, *JHEP* **04** (2015) 040, arXiv: [1410.8849](https://arxiv.org/abs/1410.8849) [[hep-ph](#)] (cit. on p. 114).
- [128] C. Anastasiou, L. Dixon, K. Melnikov and F. Petriello, *High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, *Phys. Rev. D* **69** (2004) 094008, arXiv: [hep-ph/0312266](https://arxiv.org/abs/hep-ph/0312266) (cit. on p. 114).
- [129] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126, arXiv: [0707.3088](https://arxiv.org/abs/0707.3088) [[hep-ph](#)] (cit. on p. 114).
- [130] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, arXiv: [hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146) (cit. on p. 114).
- [131] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, arXiv: [0709.2092](https://arxiv.org/abs/0709.2092) [[hep-ph](#)] (cit. on p. 114).

- [132] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043, arXiv: [1002.2581 \[hep-ph\]](#) (cit. on p. 114).
- [133] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152 (cit. on p. 114).
- [134] L. Carloni and T. Sjöstrand, *Visible effects of invisible Hidden Valley radiation*, *Journal of High Energy Physics* **2010** (2010), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP09\(2010\)105](http://dx.doi.org/10.1007/JHEP09(2010)105) (cit. on p. 114).
- [135] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *JHEP* **03** (2020) 145, arXiv: [1910.08447 \[hep-ex\]](#) (cit. on pp. 115, 188).
- [136] CMS Collaboration, *Search for narrow and broad dijet resonances in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$ and constraints on dark matter mediators and other new particles*, *JHEP* **08** (2018) 130, arXiv: [1806.00843 \[hep-ex\]](#) (cit. on pp. 115, 188).
- [137] CMS Collaboration, *Search for high mass dijet resonances with a new background prediction method in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$* , *JHEP* **05** (2020) 033, arXiv: [1911.03947 \[hep-ex\]](#) (cit. on pp. 115, 188).
- [138] W. Buttinger, ‘Background Estimation with the ABCD Method Featuring the TRooFit Toolkit’, 2018, URL: <https://api.semanticscholar.org/CorpusID:235806829> (cit. on pp. 132, 133).
- [139] ATLAS Collaboration, *Differential $t\bar{t}$ cross-section measurements using boosted top quarks in the all-hadronic final state with 139 fb^{-1} of ATLAS data*, *JHEP* **04** (2023) 080, arXiv: [2205.02817 \[hep-ex\]](#) (cit. on p. 136).
- [140] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *The European Physical Journal C* **71** (2011), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-011-1554-0> (cit. on p. 174).
- [141] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, *Transactions of the American Mathematical Society* **54** (1943) 426, ISSN: 00029947, 10886850, URL: <http://www.jstor.org/stable/1990256> (visited on 17/07/2024) (cit. on p. 179).
- [142] L. Heinrich, M. Feickert and G. Stark, *pyhf: v0.7.6*, version 0.7.6, <https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6>, URL: <https://doi.org/10.5281/zenodo.1169739> (cit. on p. 180).
- [143] L. Heinrich, M. Feickert, G. Stark and K. Cranmer, *pyhf: pure-Python implementation of HistFactory statistical models*, *Journal of Open Source Software* **6** (2021) 2823, URL: <https://doi.org/10.21105/joss.02823> (cit. on p. 180).
- [144] M. Proffitt, *pyhf-based implementation of the ABCD method for background estimation*, version 0.0.6, URL: <https://pypi.org/project/abcd-pyhf/> (cit. on p. 180).

- [145] ATLAS Collaboration,
Dark matter summary plots for s-channel, 2HDM+a, Higgs portal and Dark Higgs models,
ATL-PHYS-PUB-2024-010, 2024, URL: <https://cds.cern.ch/record/2901527>
(cit. on p. 189).
- [146] J. Alwall 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 **2014** (2014), ISSN: 1029-8479,
URL: [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079) (cit. on p. 188).

