Alma Mater Studiorum · University of Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Dijet background estimation in the hadronic $V+$jets cross section measurement

Supervisor:
Prof. Maximiliano Sioli

Co-supervisor:
Dr. Chris M. Delitzsch

Submitted by:
Martina Turchini

Academic Year 2023/2024

Abstract

Accurate background estimation is crucial for precision measurements in high-energy physics. This thesis presents a study on the estimation of the dijet background in the measurement of the hadronic vector boson plus jets ($V+$jets) cross section using Monte Carlo (MC) simulated data from the ATLAS experiment at the Large Hadron Collider (LHC).

In this analysis, hadronically decaying W and Z bosons are reconstructed as large-radius jets. Their identification is particularly challenging due to the overwhelming background from Quantum Chromodynamics (QCD) multijet processes, whose production cross sections are several orders of magnitude larger than those of the signal. To model this background, MC simulations are employed, focusing on the boosted regime where theoretical and modeling uncertainties are significant. Smooth, analytically functions are fitted to the MC-derived background distributions in control regions adjacent to the signal region. This approach allows for a flexible and robust modeling of the QCD background, which is crucial for accurate signal extraction. The results provide insights into the reliability of background estimation methods, which are critical for precision measurements and for enhancing the sensitivity of searches for new physics phenomena involving boosted vector bosons.

Acknowledgments

I would like to express my sincere gratitude to my supervisor at the University of Bologna, Prof. Sioli, for his guidance, support, and encouragement throughout my academic path, starting from my Bachelor's thesis to the completion of this Master's work.

I am equally grateful to my supervisor at TU Dortmund, Dr. Chris M. Delitzsch, for her continuous mentorship and for warmly welcoming me into her research group. The collaborative and stimulating environment she and her team created greatly enriched my research experience. In particular, I am especially thankful to Donna M. Mattern and Dr. Amartja Rej for their direct help and valuable assistance with the data analysis.

I would also like to thank all the members of the Kröninger group for their support and constructive discussions during my time in Dortmund.

Finally, I would like to acknowledge the Erasmus+ program for enabling my study experience in Germany, which was a crucial and enriching part of my academic journey.

Contents

List of Figures

List of Tables

List of Acronyms

ANN Adversarial Neural Network. 2, 3, 32, 43, 45, 50

BSM Beyond the Standard Model. 1, 4, 13

CERN Conseil Européen pour la Recherche Nucléaire. 15, 16

CR Control Region. 34, 45, 50

EWT Electroweak Theory. 9

HL-LHC High-Luminosity LHC. 18, 26

LHC Large Hadron Collider. 1, 2, 4, 12, 15, 17–19, 24, 26, 34

MC Monte Carlo. 18, 19, 29, 31, 40

QCD Quantum Chromodynamics. 1–3, 7, 11, 13, 29, 32–34, 40–42, 50

QED Quantum Electrodynamics. 7, 8

QFT Quantum Field Theory. 7, 9

SM Standard Model. 1, 2, 4–8

SR Signal Region. 2, 32, 34–38, 40

UFO Unified Flow Object. 13

Chapter 1

Introduction

The production of vector bosons (W or Z) in association with jets (V+jets) represents a key process in the phenomenology of the Standard Model (SM) at hadron colliders. In particular, measurements of the cross section for hadronically decaying vector bosons provide a stringent test of perturbative Quantum Chromodynamics (QCD), as they involve both electroweak boson production and QCD-induced jet radiation, making them sensitive to higher-order corrections and parton shower modeling [1]. Furthermore, V+jets processes represent a significant background in several SM measurements. Accurate modeling of these processes is therefore crucial not only for improving the accuracy of Higgs property measurements but also for controlling uncertainties in multi-boson and top-quark associated production processes. For instance, they contribute substantially to backgrounds in studies of Higgs boson production—such as in the $H \to WW^*$ [2] decay channel—where the final states include jets and missing transverse momentum. Moreover, since V+jets processes also constitute a dominant background in searches for heavy resonances and Beyond the Standard Model (BSM) phenomena, achieving a precise understanding of their behaviour is essential.

At the Large Hadron Collider (LHC), proton–proton (pp) collisions at centre-of-mass energies of up to 13.6 TeV enable the production of vector bosons with high transverse momentum. When such bosons decay hadronically, the resulting quarks hadronize into collimated sprays of particles, referred to as jets. In the highly boosted regime, the two quarks from the decay of a W or Z boson may be reconstructed as a single, large-radius jet. The identification of these so-called "W/Z jets" plays a critical role in extending the sensitivity of a variety of different searches for BSM physics, for instance Ref. [3, 4].

Thus far, research has predominantly concentrated on the leptonic channels of vector bosons, such as $W \to e\nu$, $W \to \mu\nu$, and $Z \to \ell\ell$, where ℓ denotes an electron, muon, or tau, for example Ref. [5]. This preference persists despite the smaller branching ratios of leptonic modes compared to hadronic decays, as reported by the Particle Data Group (PDG) [6]. Moreover, these leptonic decay modes yield final states that are relatively easier to identify and reconstruct within particle detectors. The presence of high-energy charged leptons provides clear experimental signatures, enabling the selection of signal events with comparatively low background contamination. This facilitates the accurate determination of fundamental parameters, including coupling constants and branching ratios. In contrast to leptonic final states, hadronic decays of vector bosons—where the bosons decay into jets formed from quarks and gluons—pose considerable experimental challenges [7].

A major experimental challenge in the hadronic V+jets final state is the overwhelming presence of background from QCD multijet production, which originates from strong interaction processes and features significantly larger cross sections. These QCD-induced jets can mimic the topology of hadronic V+jets events, making event-by-event discrimination extremely difficult. Consequently, a reliable estimation of the QCD background is essential to extract meaningful physical results and ensure a correct interpretation of the data.

One of the experiments dedicated to studying V+jets events is ATLAS, a general-purpose detector at the LHC designed to accurately measure a wide variety of final states. Its sophisticated calorimetry and tracking systems enable the reconstruction of jets with different algorithms and radius parameters, allowing for detailed investigations of both standard and boosted event topologies [8].

This thesis focuses on the estimation of the dijet background in the context of a V+jets cross section measurement, where the vector boson decays hadronically and is reconstructed as a large-radius jet. The analysis is based on simulated data samples produced for the ATLAS experiment at the LHC and involves fitting the background distribution using smooth analytical functions in dedicated control regions adjacent to the signal peak. Further studies investigate the use of machine learning techniques—particularly Adversarial Neural Network (ANN) [9]—to suppress the overwhelming QCD background and to study their impact on the shape of the background distribution. This allows for a more accurate estimation of the background in the Signal Region (SR) and helps quantify potential biases introduced by such methods.

The structure of this thesis is as follows. Chapter 1 will introduce the theoretical framework fundamental to this work. An overview of the SM will be provided, including the fundamental interactions it describes. Relevant kinematic variables used in high-energy collisions will then be discussed, followed by a description of jet formation mechanisms.

Chapter 2 provides an overview of the experimental setup, beginning with the general features of the LHC and proceeding to a detailed description of the ATLAS detector and its principal components. Key collision parameters essential to the study will also be introduced.

Chapter 3 describes the simulation samples used throughout the analysis and outlines the procedure for deriving signal and background distributions. In addition, it introduces the ANN classifier, which will be employed in Chapter 5.

Chapter 4 will then focus on the methods used to estimate the QCD background. The sideband fitting procedure will be introduced, followed by a discussion of fit performance and model comparisons.

Chapter 5 will present the studies conducted using an ANN classifier applied to the background distribution. This includes the sideband fitting procedure, signal-plus-background fits, and a series of validation tests—such as spurious signal and signal injection tests—designed to assess the robustness and stability of the results.

Chapter 2

The Standard Model

The Standard Model of Particle Physics, developed in the early 1970s, represents one of the greatest achievements of modern science, providing a clear and comprehensive description of fundamental particles and their interactions—also called forces. In 2012, the first observation of the Higgs boson at the LHC marked the experimental confirmation of the last undiscovered particle predicted by the SM, which had been theoretically complete for decades [10].

The SM encapsulates our understanding of the fundamental structure of matter, offering profound insights into the elementary building blocks of the universe [11].

Since its first formulation, the SM has been supported by extensive experimental evidence. Although it is a remarkably successful theory, it still leaves many fundamental questions unanswered, such as:

- the origin of neutrinos masses;
- dark mass and energy existence explanation;
- the explanation of the dominance of baryonic matter over the anti-baryonic one in the Universe.

Moreover, the formulation of the SM does not incorporate the gravitational interaction. Despite these limitations, experimental results obtained over the last decade seem to point toward the existence of physics BSM.

In the following, the elementary particles will be introduced in Section 2.1, followed by a discussion of the the fundamental forces as described by the SM in Section 2.2. Section 2.3 will then cover the relevant kinematic variables used in high-energy collisions. Finally, Section 2.4 will present the concept of jets and highlight their relevance in high-energy physics, as they play a central role in the analysis conducted in this work.

2.1 Particles

According to the SM, fundamental particles are divided in two main families: the matter constituents (fermions) and those responsible for mediating the fundamental interactions (bosons).

A schematic representation of the fundamental particles of the SM and their main properties can be found in Figure 2.1.

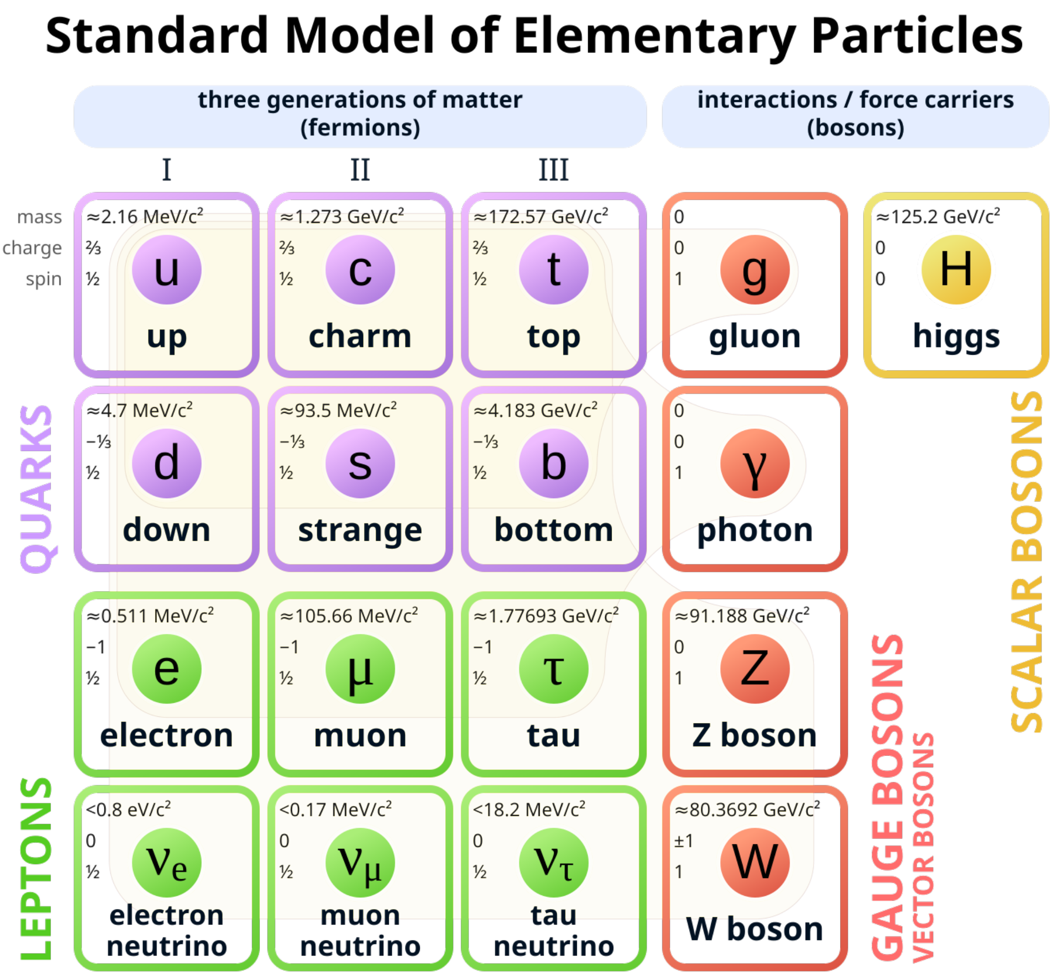


Figure 2.1: Fundamental particles of the SM with their main properties [12].

Fermionic particles are characterized with spin 1/2 and behave according to the Fermi-Dirac statistics, and therefore to Pauli's exclusion Principle [13]. Fermions are divided in two families according to the interaction they "respond" to and their charge:

leptons and quarks. The former are subjected to the electromagnetic (EM) and weak interactions, while the latter are also subject to the strong interaction.

In the SM, fermions are composed of six leptons (electrons, muons, taus and three neutrinos) and six quarks (up, charm, top, down, strange, bottom) divided in three generations according to their weak interaction properties:

$$\text{Leptons:} \quad \begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{2.1}$$

$$\text{Quarks:} \quad \begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix} \tag{2.2}$$

Each of the fermions listed in Equation 2.1 and 2.2 has a corresponding antiparticle with identical mass, spin and half-life, but opposite internal charge.

These internal charges, or quantum numbers, are fundamental properties that determine how particles interact through the fundamental forces. In the SM, particles are characterized by several quantum numbers, including the electric charge Q, color charge— introduced in Section 2.2.1—weak isospin T_3, and hypercharge Y_W—both discussed in Section 2.2.3—as well as lepton and baryon numbers.

These quantities are conserved in most interactions and are essential for classifying particles and predicting allowed processes. The electric charge Q is expressed in units of the elementary charge e[1]. Electrons, muons, taus, and all quarks carry a non-zero electric charge, while neutrinos are electrically neutral.

The baryon number B is a conserved quantum number that prevents the transformation of baryons into non-baryonic particles. Quarks are assigned a baryon number of $+\frac{1}{3}$, while antiquarks carry $-\frac{1}{3}$. Since baryons—such as the proton— are composed of three quarks, they have $B = +1$, whereas antibaryons have $B = -1$. Mesons, which consist of a quark-antiquark pair, carry a baryon number of zero. Similarly, leptons, which do not interact via the strong interaction, also have a baryon number of zero.

The lepton number L conservation forbids the transformation of leptons into bosons or baryons. There are three lepton numbers, one for each lepton family: the electronic lepton number L_e, the muonic lepton number L_μ, and the tauonic lepton number L_τ. For instance, $L_e = +1$ is assigned to the electron e^- and its neutrino ν_e, while their antiparticles e^+ and $\bar{\nu}_e$ have $L_e = -1$. Leptons of the other two families have $L_e = 0$. The same logic applies for L_μ and L_τ.

The baryon and lepton numbers are strictly conserved in all known physical processes, regardless of the type of interaction involved. Other quantum numbers—such as isospin or parity—are conserved only under specific forces, like the electromagnetic or strong interaction, but not necessarily under weak interactions [14].

[1] $e \sim 1.6022 \times 10^{-19}$ C

2.2 Fundamental Interactions

The SM describes interactions between particles as a Quantum Field Theory (QFT), corresponding to the exchange of characteristic bosons [11]. These bosons are the mediators of the fundamental forces; they have integer spin and therefore obey Bose-Einstein statistics. The SM predicts several types of gauge bosons:

- Photons (γ): They are the carriers of the electromagnetic force, described by Quantum Electrodynamics (QED). They are massless, electrically neutral boson with spin 1.
- W^+, W^-, and Z bosons: They mediate the weak interaction. The W^+ and W^- bosons have positive and negative electric charge, respectively, while the Z boson is neutral. All three of them are massive.
- Eight gluons: They are the mediators of the strong interaction. They are massless but carry color charge, which causes them to self-interact.

In addition to these, there is the scalar Higgs boson, which is massive, has zero spin, and carries neither electric nor color charge. It mediates the Higgs field, which interacts with other elementary particles through the Brout-Englert-Higgs (BEH) mechanism, providing them with mass [15].

2.2.1 Quantum Chromodynamics (QCD)

QCD is a non-Abelian QFT based on the SU(3) gauge symmetry group. It describes the strong interaction, which governs the behavior of quarks and gluons and is responsible for the binding of nucleons within atomic nuclei. The theory is mediated by eight massless gluons, each corresponding to one of the eight generators of SU(3). QCD introduces three conserved "color" charges—red, blue, and green—which label the orthogonal states in SU(3) color space. Only particles that possess a non-zero color charge can couple to gluons and interact via the strong interaction.

An important consequence of gluons themselves carrying color charge is the emergence of two key phenomena characteristic of the strong interaction: confinement and asymptotic freedom. Confinement refers to the fact that at low energies, only color-neutral particles can exist; in nature, quarks are never observed in isolation but always in bound (hadronic) states with a net color charge of zero ("colorless") [16, 17]. Asymptotic freedom, on the other hand, describes how the strength of the strong force decreases at high energy scales, allowing quarks to behave as if they were nearly free when probed at very short distances.

2.2.2 Quantum Electrodynamics (QED)

The electromagnetic (EM) interaction is the only long-range fundamental force described by the Standard Model. Well known in classical physics through Maxwell's equations, it is formulated at the quantum level by QED—the relativistic quantum field theory that describes the interaction between light and matter. QED provides a comprehensive framework for understanding how electrically charged particles interact via the electromagnetic field and is one of the most precisely tested theories in physics.

The photon's zero mass allows the electromagnetic force to have an infinite range, while its lack of electric charge prevents photons from self-interacting. Consequently, electromagnetic processes involving the emission or absorption of photons always conserve electric charge.

As discussed in Section 2.1, all SM particles interact electromagnetically, with the exception of neutrinos, which are electrically neutral and thus unaffected by this force. The strength of the electromagnetic interaction is characterized by the fine-structure constant:

$$\alpha_{em} = \frac{e^2}{4\pi\epsilon_0\hbar c} \sim \frac{1}{137}$$

where ϵ_0 is the vacuum permittivity and $\hbar$[2] is the reduced Planck constant. This dimensionless constant provides a measure of the coupling strength between electrically charged particles and the electromagnetic field [14].

2.2.3 Electroweak Theory (EWT)

The electroweak interaction, or electroweak force, is the unified theoretical framework that combines two of the four fundamental forces in nature: electromagnetism and the weak nuclear force. While these forces appear very different at low energies, they are understood to be two manifestations of a single interaction at higher energy scales. According to the electroweak theory, above the unification energy (approximately 246 GeV[3]), the electromagnetic and weak forces merge into a single force [17].

The first theoretical attempt to describe the weak interaction was proposed by Enrico Fermi to explain beta decay[4] [14]. The weak force is mediated by the massive gauge

[2] $\hbar = h/2\pi \sim 1.054 \times 10^{-34}$ J · s [18].

[3] This corresponds to the vacuum expectation value of the Higgs field $v = \left(\sqrt{2}G_F\right)^{-1/2} \approx 246$ GeV, where G_F is the Fermi constant.

[4] Beta decay is a process in which a neutron decays into a proton, an electron, and an electron neutrino:

$$n \to p + e^- + \bar{\nu}_e$$

The continuous energy spectrum of the emitted electron was the first indication that beta decay is a three-body decay, suggesting the existence of an additional, invisible particle—later identified as the neutrino.

bosons W^+, W^-, and Z^0. Their non-zero mass explains the short-range nature of the weak interaction, in contrast to the infinite range of the electromagnetic force. According to QFT, the mass of a force carrier determines the range of the force it mediates: massive bosons lead to exponentially suppressed interactions at long distances, as described by the Yukawa potential [17].

A distinctive feature of the weak interaction is that it violates several symmetries that are conserved in both electromagnetic and strong interactions. These include parity (P) violation[5], charge conjugation (C) violation, and flavor non-conservation. In nature, the only stable hadrons observed are the proton and the neutron—the latter being stable only within the atomic nucleus.

The weak interaction is also responsible for quark flavor mixing, a phenomenon described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix [20, 21], which encodes the probability amplitudes of transitions between different quark flavors during weak processes:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} \tag{2.3}$$

In Equation 2.3, the flavor eigenstates, denoted with primed letters, are related to the mass eigenstates through the matrix elements V_{ij}.

It was the interdependence between electric charge and weak charge that led Glashow, Salam, and Weinberg to develop a unified theory of the electromagnetic and weak interactions: the Electroweak Theory (EWT). The charge associated with this interaction is called the weak hypercharge Y_W, defined via the Gell-Mann–Nishijima relation:

$$Y_W = 2(Q - T_3) \tag{2.4}$$

where Q is the electric charge and T_3 is the third component of the weak isospin. The values of the weak hypercharge Y_W for the various fermions, along with their electric charge Q and weak isospin T_3, are shown in Table 2.1. For quarks and antiquarks, the first value corresponds to up-type quarks, while the second refers to down-type quarks.

2.3 Relevant Kinematic Variables

In high-energy collisions, kinematic variables play a crucial role in characterizing the interaction, providing insights into the energy, momentum, and angular distribution of the resulting particles. This section describes some of the kinematic variables relevant for the purposes of this thesis:

[5] Parity violation was experimentally confirmed in 1956 by physicist Chien-Shiung Wu, through her groundbreaking work on the beta decay of cobalt-60 nuclei. Her results provided direct evidence that the weak interaction does not conserve parity symmetry [19].

Table 2.1: Quantum numbers Q, T_3, and Y_W for different fermion types.

	Leptons	**Antileptons**	**Quarks**	**Antiquarks**
Q	-1	$+1$	$+\frac{2}{3}, -\frac{1}{3}$	$-\frac{2}{3}, +\frac{1}{3}$
T_3	$-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}, -\frac{1}{2}$	$-\frac{1}{2}, +\frac{1}{2}$
Y_W	-1	$+1$	$+\frac{1}{3}$	$-\frac{1}{3}$

- **Invariant mass** m

  $$m^2 = E_{\text{tot}}^2 - |\vec{p}_{\text{tot}}|^2$$

  This quantity is obtained from the relativistic energy-momentum relation, where E_{tot} and $\vec{p}_{\text{tot}}$ are respectively the total energy and total 4-momentum vector of the system (in natural units, $c = 1$). In high-energy physics, the invariant mass is a key observable, often used to search for new particles or resonances.

- **Transverse momentum** p_T

  $$p_T = \sqrt{p_x^2 + p_y^2}$$

  where p_x and p_y are the momentum components orthogonal z-axis along the beam pipe [22].

- **Rapidity** y

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

  where p_z is the momentum component along the beam axis and E is the particle's energy.

- **Pseudorapidity** η is used to describe the angle between a particle and the beam axis:

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

  where θ is the angle between the particle's momentum $\vec{p}$ and the beam axis.

  Rapidity depends on the particle's energy and longitudinal momentum and is Lorentz invariant under boosts[6] along the beam axis, while pseudorapidity depends only on the emission angle, reflecting the detector geometry. In the ultra-relativistic limit $(E \gg m)$ η approximates y.

[6] A Lorentz boost is a Lorentz transformation without a rotation [23].

2.4 Jets

As discussed in Section 2.2.1, particles that carry color charge—such as quarks and gluons, collectively known as partons—cannot exist in isolation in nature. Instead, they are bound into color-neutral particles due to the phenomenon of confinement. In high-energy pp collisions, the colored partons inside the protons interact and exchange color charge. To satisfy confinement, these color-charged fragments undergo further interactions, generating additional partons that eventually hadronize into color-neutral particles. The result is a collimated stream of particles moving in approximately the same direction, forming a narrow, localized structure known as a jet.

Jets manifest in particle detectors as localized clusters of energy and serve as experimental signatures of quarks or gluons produced in high-energy collisions. Their reconstruction is essential for interpreting the underlying physics processes and is carried out using dedicated algorithms that cluster detector signals into jets. Depending on the chosen reconstruction approach, various types of inputs can be employed, including energy deposits in the calorimeter, charged particle tracks from the inner detector, or a combination of both.

After reconstructing inputs, sequential recombination algorithms are commonly used. These algorithms define a set of distances between input objects, iteratively combining them until all jets are formed. A crucial requirement for any jet clustering algorithm is that it satisfies infrared and collinear safety (ICS), meaning the algorithm's output should remain unchanged under the emission of a soft particle or the collinear splitting of a particle. These properties ensure the theoretical robustness of jet definitions and allow for meaningful comparisons between experimental data and perturbative QCD predictions. Several algorithms satisfy these requirements, including the k_t algorithm [24], the Cambridge/Aachen (C/A) algorithm [25, 26], and the anti-k_t algorithm [27]. Each of these algorithms is characterized by a parameter p, which determines the distance metric and, consequently the clustering behavior. The clustering procedure involves defining distances d_{ij} between entities i and j, and d_{iB} between entity i and the beam B. At each step, the smallest distance is identified: if it is d_{ij}, entities i and j are recombined; if it is d_{iB}, entity i is declared a jet and removed. This repeats until no entities remain. The distance is defined as:

$$d_{ij} = \min\left(p_{Ti}^{2p}, p_{Tj}^{2p}\right)\frac{\Delta_{ij}^2}{R^2}, \quad d_{iB} = k_{ti}^{2p}, \tag{2.5}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, with p_{Ti}, y_i, and ϕ_i the transverse momentum, rapidity, and azimuth of particle i. The parameter R sets the effective clustering radius in rapidity–azimuth space. It determines how close particles must be to be merged into the same jet and thus controls the angular extent and capture area of the resulting jet. The exponent p determines the behavior of the algorithm and defines the specific clustering scheme.

For $p = 1$ the distance measure corresponds to the inclusive k_t algorithm; for $p = 0$ it reduces to the inclusive Cambridge/Aachen algorithm. The anti-k_t algorithm, corresponding to $p = -1$, is widely adopted at the LHC due to its tendency to produce geometrically regular, conical jets that are less sensitive to soft radiation, making them easier to calibrate and cluster into jets [27]. Consequently, the anti-k_t algorithm was employed in this analysis to reconstruct jets.

In the anti-k_t algorithm, since distances scale as the inverse square of transverse momentum in the anti-k_t algorithm, particles with higher p_t have smaller d_{iB}, and thus are clustered earlier in the algorithm.

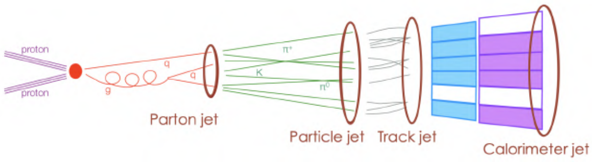


Figure 2.2: Sketch of pp collision and resulting jet development [28].

Jets are powerful tools for probing the underlying physics, but they remain approximations of the actual partonic processes. Depending on the analysis goals, different jet definitions may be more appropriate. At ATLAS, small-radius (small-R) jets—defined with a radius parameter $R = 0.4$—are typically used to study quarks and gluons, whereas large-radius (large-R) jets—with $R = 1.0$—are particularly useful for identifying hadronically decaying heavy particles, such as W and Z bosons. This extended radius is essential for capturing all decay products within a single jet, as the angular separation ΔR between the decay products of a massive particle χ scales approximately as

$$\Delta R \sim \frac{2m_\chi}{p_{T,\chi}}.$$

Thus, for jets with $p_T > 200$ GeV, most of the decay products from hadronically decaying W and Z bosons are expected to be contained within a jet radius of $R = 1.0$.

In both cases, jets require proper calibration to accurately reflect the event's kinematics, after which tagging techniques can be applied to infer their likely particle origin [29].

Section 2.4.1 provides a more detailed description of large-R jets.

2.4.1 Large Radius Jets

A boosted particle refers to a particle χ produced with $p_T > m_\chi$. In the laboratory frame, the decay of such a particle (e.g., jets) results in its products being emitted within a narrow cone, due to Lorentz boosting [23] compressing the decay angles. This collimation makes standard jet reconstruction techniques insufficient, as multiple partons may merge into a single, large-R jet [30].

To identify the origin of these boosted particles, advanced jet substructure techniques are employed. These are especially important in identifying merged jets from highly boosted W/Z bosons, as illustrated in Figure 2.3, enhancing the sensitivity of heavy resonances and potential BSM signatures.

In ATLAS, the reconstruction of large-R jets relies on a specific type of input objects called Unified Flow Objects (UFOs), designed to capture the broader and more complex structure of such jets. UFOs combine information from both the inner tracking detectors and the calorimeters, effectively integrating the strengths of both subsystems. Tracking detectors provide excellent angular resolution and superior moment resolution at low p_T, making them particularly effective for resolving fine jet substructure. However, their momentum resolution degrades at high p_T, where the curvature of charged particle tracks becomes harder to measure precisely in the magnetic field. In contrast, calorimeter-based measurements become increasingly effective at higher p_T. By combining both types of measurements, UFOs enhance jet reconstruction across a wide kinematic range and significantly improve the ability to identify the internal structure of jets arising from boosted hadronic decays, such as those of W/Z [31, 32, 33, 8].

In the following analysis, UFO jets reconstructed with a radius parameter $R = 1.0$ were used, combined with soft drop declustering [34] to reduce contamination from pile-up and soft radiation. This procedure improves the accuracy of jet property measurements and enhances tagging performance[7].

Figure 2.3 also illustrates the distinction between signal jets and QCD background jets. Signal jets typically arise from the hadronic decays of heavy, boosted resonances such as W or Z bosons. Due to Lorentz boosting—denoted by $\Lambda_z(\beta)$—their decay products become highly collimated and are often contained within a single large-R jet. These jets tend to display a relatively simple and symmetric substructure. In contrast, QCD jets originate from the fragmentation of quarks and gluons via strong interactions. They generally show a broader and more irregular energy distribution, with a more complex and less correlated substructure. The substructure of jets can be used to distinguish between signal jets and background jets, but this task becomes more challenging at high p_T when the decay products become even more collimated and overlapping [36].

A key distinguishing feature is the jet mass: jets originating from W/Z decays typically have a mass close to the boson mass, around 80 GeV for W to 90 GeV for Z, whereas quark- or gluon-initiated jets tend to have a lower, but non-zero mass, arising from wide-angle radiation in the parton shower. This mass difference provides a powerful discriminator between signal and background [37].

[7] Tagging performance refers to the ability to correctly identify the jet's origin, distinguishing jets from hadronic decays of boosted massive particles (such as W or Z) from background jets [35].

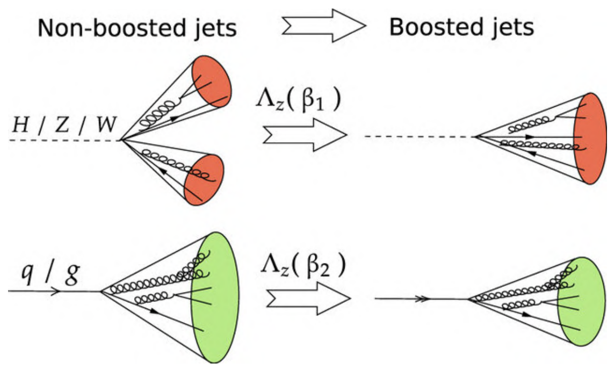


Figure 2.3: Illustration of non-boosted (left) and boosted (right) jets in high-energy particle collisions. In the upper part of the image, a heavy particle (such as a Higgs, W, or Z boson) decays into two partons that give rise to jets. In the non-boosted scenario, the decay products are well-separated, resulting in two distinct small-R jets. In the boosted regime (right), due to a large Lorentz boost $\Lambda_z(\beta)$, the decay products are collimated into a single, large-R jet. In the lower part of the image, quark- or gluon-initiated jets are shown under similar non-boosted and boosted conditions [36].

Chapter 3

The ATLAS experiment at LHC

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator. Located at the *Conseil Européen pour la Recherche Nucléaire (CERN)* in Geneva, it first started up on 10 September 2008, and remains the latest addition to CERN's accelerator complex [38].

The accelerator complex consists of a sequence of devices that progressively accelerate particles—typically protons—to higher energies, as well as detectors that record the products of the collisions between these accelerated particles. The layout of the complex is shown in Figure 3.1 and includes LINAC2, a linear accelerator, followed by four circular accelerators: the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS), which inject the particle beams directly into the LHC. At each stage of the acceleration process, the beam energy is progressively increased until it reaches the maximum value achieved in the LHC, namely 6.8 TeV per beam. This corresponds to a center-of-mass energy of $\sqrt{s} = 13.6$ TeV, attained during the most recent data-taking period, known as Run 3, which is presently ongoing [39].

Looking ahead, CERN is exploring the next generation of particle accelerators through the Future Circular Collider (FCC) study. The FCC aims to significantly extend the physics research beyond the LHC by first implementing a high-luminosity electron-positron collider (FCC-ee) for precision studies, followed by a *pp* collider (FCC-hh), capable of reaching collision energies up to 100 TeV, far beyond LHC. Located in a new 90.7 km circular tunnel, this facility would enable physicist to investigate key open questions in fundamental physics, such as the nature of dark matter and the role of the Higgs boson in the evolution of the universe [40].

3.1 Large Hadron Collider (LHC)

LHC is the largest particle accelerator ever built. It is located on the site of the former Large Electron-Positron Collider (LEP) and serves as the final stage of proton accelera-

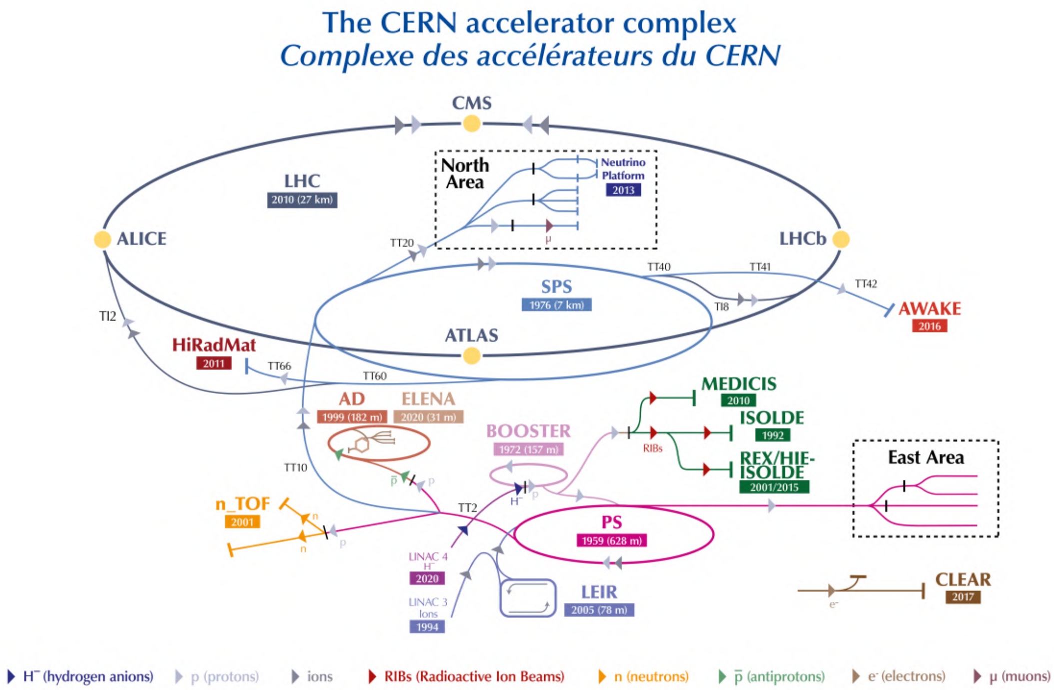


Figure 3.1: The CERN accelerator complex showing the different accelerators and detectors [41].

tion in CERN's accelerator complex. It is an circular ring with a circumference of 27 km, located underground at depths ranging from 50 to 175 meters. Inside the accelerator, two high-energy particle beams travel in opposite direction at velocities close to the speed of light before colliding with each other. The separated tubes in which the beams travel must be maintained at an ultra-high vacuum condition, corresponding to the extremely low pressure of about 10^{-10} torr—comparable to the pressure in the Interstellar Medium (ISM)—to prevent collisions between the beam particles and the gas molecules inside the tube [42].

The beams are guided by an intense magnetic field provided by superconducting magnets. To operate efficiently and without significant energy loss, these magnets are kept at a temperature of 1.9 K using a helium-based refrigeration system [43]. After reaching the maximum energy value, the two counter-rotating beams are brought to collision at four interaction points, corresponding to the four main experiments:

- **ATLAS (*A Toroidal LHC Apparatus*)**: It is an multi-purpose detector dedicated to investigating a wide range of physical processes, from precision measurements of Standard Model bosons—such as the Higgs boson—to the search for particles that could constitute dark matter, through collisions between protons or heavy ions [44].

- **CMS (*Compact Muon Solenoid*)**: It serves the same purpose as ATLAS but employs different detection technologies [45].
- **ALICE (*A Large Ion Collider Experiment*)**: It focuses on heavy-ion physics, investigating conditions of extremely high energy density in which a state of matter known as quark-gluon plasma is formed [46].
- **LHCb (*Large Hadron Collider beauty*)**: It is dedicated to studying the beauty quark, with the goal of exploring the differences between matter and antimatter [47].

3.1.1 Beam Structure and Collision Parameters at the LHC

Each of the two proton beams circulating in the LHC is composed of 2808 bunches, with each bunch containing approximately 10^{11} protons. These bunches are spaced 25 ns apart, resulting in a bunch crossing rate of 40 MHz. However, due to operational gaps, the average crossing rate is about 30 MHz [48].

Two key parameters define the performance of a particle accelerator: the center-of-mass energy ($\sqrt{s}$) and the instantaneous luminosity (L).

According to Einstein's mass-energy equivalence principle, $E = mc^2$, a portion of the energy released during a collision can be transformed into mass, resulting in the creation of new particles. Therefore, the greater the collision energy, the more massive the particle that can be produced. Additionally, the Heisenberg uncertainty principle, expressed as $\Delta x \cdot \Delta p \simeq \frac{\hbar}{2}$, implies that probing smaller distance scales requires higher momenta, and thus higher energies [14].

The invariant mass squared (s) of a two-particle system is given by the square of the sum of their four-momenta:

$$s = (p_1^* + p_2^*)^2 = E_{\text{cm}}^2 \tag{3.1}$$

where p_1^* and p_2^* are the four-momenta of the colliding particles, and E_{cm} is the total energy in the center-of-mass frame. The LHC achieved a center-of-mass energy of $\sqrt{s} =$ 13.6 TeV during Run 3, enabling the production of massive particles and the exploration of phenomena at unprecedented energy scales [49].

The instantaneous luminosity L is a measure of the collision rate per unit area and time. It is related to the event rate $\dot{N}$ for a process with cross-section σ by:

$$\dot{N} = \sigma \cdot L \tag{3.2}$$

For a circular collider like the LHC, the instantaneous luminosity can be expressed in terms of beam parameters as:

$$L = \frac{1}{4\pi} \frac{n_b N_1 N_2 f_{\text{rev}}}{\epsilon_N \beta^*} \tag{3.3}$$

where n_b is the number of bunches per beam, N_1 and N_2 are the number of particles per bunch in each beam, f_{rev} is the revolution frequency, ϵ_N is the normalized transverse emittance, and β^* is the beta function at the interaction point [50].

The LHC was designed to reach an instantaneous luminosity of $1 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. This target was achieved and even surpassed during Run 2, with peak luminosities reaching up to $2 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ [51].

3.1.2 LHC Run Conditions and Pile-Up Environment

The LHC's data-taking activities are structured into operational phases known as runs, which are alternated with extended periods called long shutdowns. These shutdowns are dedicated to major upgrades of the accelerator complex and the detectors. During Run 1, between 2011 and 2012, pp collisions occurred at the centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ in 2011 and $\sqrt{s} = 8\,\text{TeV}$ in 2012. Over this period, the LHC delivered an integrated luminosity of $22.8\,\text{fb}^{-1}$, with $21.3\,\text{fb}^{-1}$ recorded by the ATLAS detector and $20.3\,\text{fb}^{-1}$ certified as suitable for physics analyses [52].

In Run 2, from 2015 to 2018, the collisions energy increased to $\sqrt{s} = 13\,\text{TeV}$. The LHC delivered an integrated luminosity of $165\,\text{fb}^{-1}$ during this phase, of which $147\,\text{fb}^{-1}$ were recorded by the ATLAS detector and $139\,\text{fb}^{-1}$ were certified as suitable for physics analyses [53]. The difference between delivered and recorded luminosity is due primarily to inefficiencies in the data acquisition system and the startup of the detectors.

Following Long Shutdown 2 (2019–2021), the LHC resumed operations in 2022 with Run 3, which is scheduled to continue through June 2026 [54]. Run 3 operates at 13.6 TeV [55] and is designed to accumulate even higher luminosity in preparation for the upcoming High-Luminosity LHC (HL-LHC) phase.

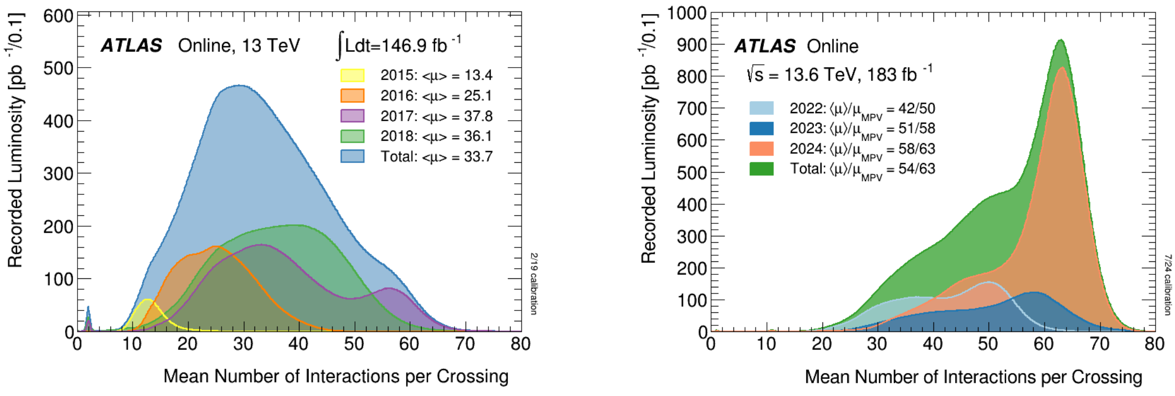
In high-luminosity conditions, multiple proton–proton interactions can occur in the same bunch crossing, a phenomenon known as pile-up. Pile-up is measured by the mean number of interactions per bunch crossing.

Figure 3.2 shows the luminosity-weighted distribution of the mean number of interactions per bunch crossing ($\langle\mu\rangle$) for Run 2 and Run 3, where $\langle\mu\rangle$ is derived from the instantaneous per-bunch luminosity using the relation

$$\mu = \frac{L_{\text{bunch}} \cdot \sigma_{\text{inel}}}{f_r},$$

with $\sigma_{\text{inel}} = 80\,\text{mb}$ and f_r the LHC revolution frequency [53]. The mean number of interactions per bunch crossing corresponds to the mean of the Poisson distribution describing the number of interactions per crossing, computed separately for each colliding bunch pair [55].

To ensure Monte Carlo (MC) simulations accurately reflect data, pile-up reweighting is applied to correct the MC pile-up profile to match that observed in data, as the true



(a) (b)

Figure 3.2: The distributions of the luminosity recorded by the ATLAS experiment as a function of the mean number of interactions per bunch crossing are shown for the entire Run 2 period and its individual years (left) [53], as well as for Run 3 data (right) [55].

profile may only be known after data-taking and can vary depending on the analysis conditions [56].

For this thesis, simulated MC events were generated under Run 2 conditions, including realistic pile-up modeling.

3.2 ATLAS Experiment

ATLAS is a general purpose detector whose goal is to cover a wide range of physical processes, from precision measurement of the SM, to search for events beyond the SM [48]. The ATLAS experiment is installed in its experimental cavern at Point 1 of the LHC, corresponding to one of the four interaction points [51].

ATLAS is a cylindrical detector measuring 46 meters in length and 25 meters in diameter, positioned 100 meters underground and weighing approximately 7,000 tonnes. The detector, as shown in Figure 3.3, is composed of multiple subsystems arranged in concentric layers around the collision point, designed to record the trajectories, momenta, and energies of particles in order to identify them individually.

To ensure precise momentum measurements, ATLAS employs a system of magnets that bend the paths of charged particles. While the detector observes over one billion interactions per second, only about one in a million is selected as potentially interesting and retained for further analysis.

ATLAS consist of four principal components, arranged in layers around the interaction point:

- **Magnet system**: responsible for deflecting the trajectories of charged particles;

- **Inner detector**: reconstructs interaction vertices and tracks the motion of charged particles;
- **Calorimeters**: measure the direction and energy of electrons, photons, and hadrons;
- **Muon spectometer**: identifies muons and determines their momentum.

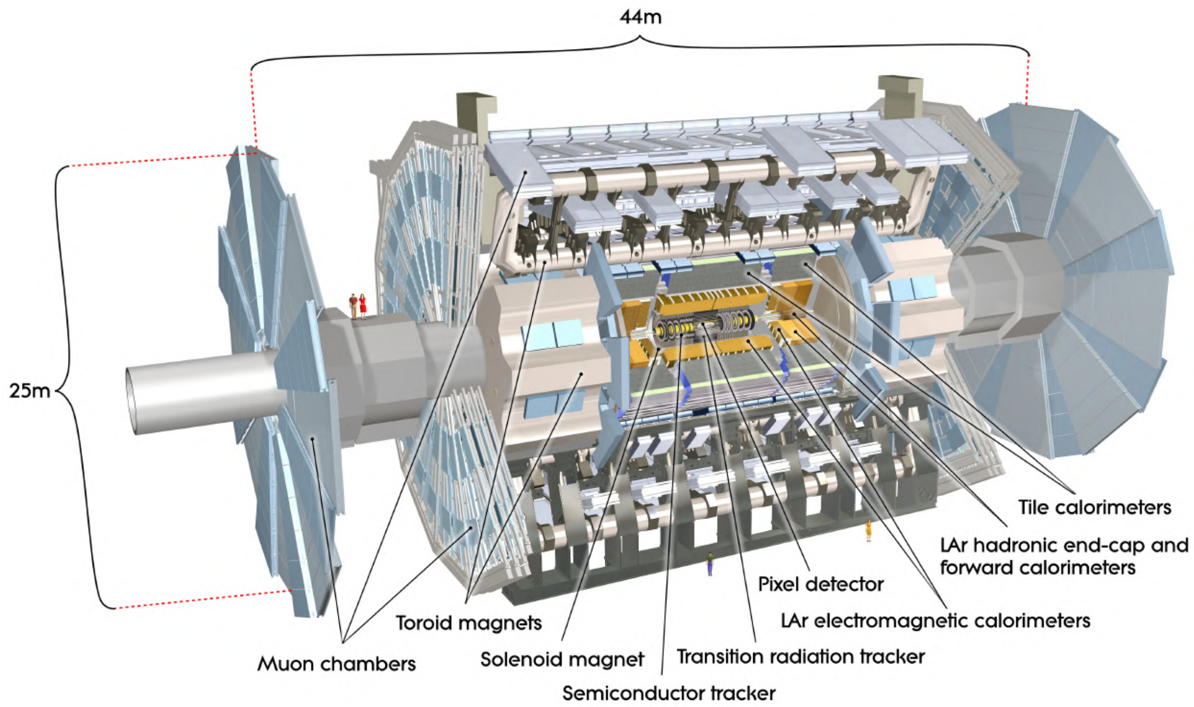


Figure 3.3: Cut-away view of the whole ATLAS detector with its main components [57].

The ATLAS coordinate system is based on polar coordinates, with the origin located at the nominal interaction point of the beams. The z-axis is defined along the beam direction, while the x–y plane is perpendicular to it. The polar angle (θ) is defined with respect to the beam axis, and the azimuthal angle (ϕ) is measured around this axis.

3.2.1 Magnet System

The magnet system is designed to measure the momentum and electric charge of particles produced from the proton-proton collision by analyzing the curvature of their trajectories. ATLAS employs two types of superconducting magnet systems—solenoidal and toroidal—which bend the path of charged particles. The magnetic field lines produced by the solenoidal and toroidal magnets are shown in Figure 3.4. These systems are maintained at a temperature of 4.5 K through a liquid helium cooling circuit.

The main components of the magnet system are the Central Solenoid Magnet, providing an axial magnetic field of 2 T for the inner tracker, the Barrel Toroid, and the End-cap Toroids together generating a toroidal magnetic field of 1 T [58].

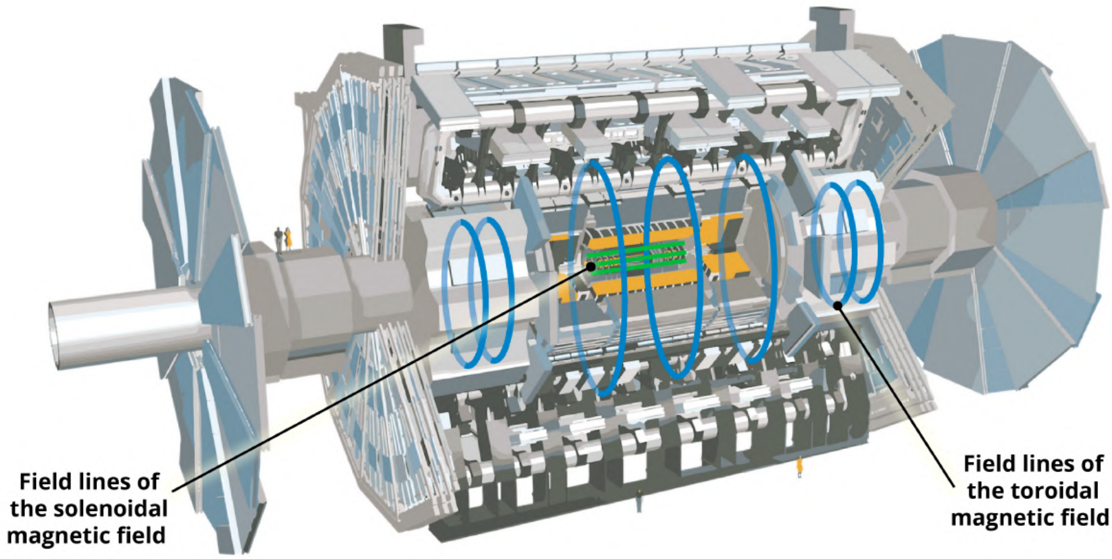


Figure 3.4: ATLAS solenoidal and toroidal magnetic system [59].

3.2.2 Inner Detector

The ATLAS Inner Detector (ID) is a compact and highly sensitive cylindrical tracking system, 6.2 m long and 2.1 m in diameter, covering the pseudorapidity range $|\eta| < 2.5$. Positioned between the beam pipe and the calorimeters, and immersed in the 2 T solenoidal magnetic field, the ID is designed for precise reconstruction of charged-particle tracks, momentum measurement, and identification of both primary (*pp* interaction) and secondary (decay) vertices. Charged particles crossing the ID deposit ionization energy, which is recorded to reconstruct their trajectories. The ID enables efficient pattern recognition, excellent momentum resolution, and provides electron identification over $|\eta| <2.0$ for energies between 0.5 and 150 GeV [60]. A cut-away view of the detector, with its main components, is visible in Figure 3.5.

The ID consists of four independent, but complementary sub-detectors, shown in Figure 3.6:

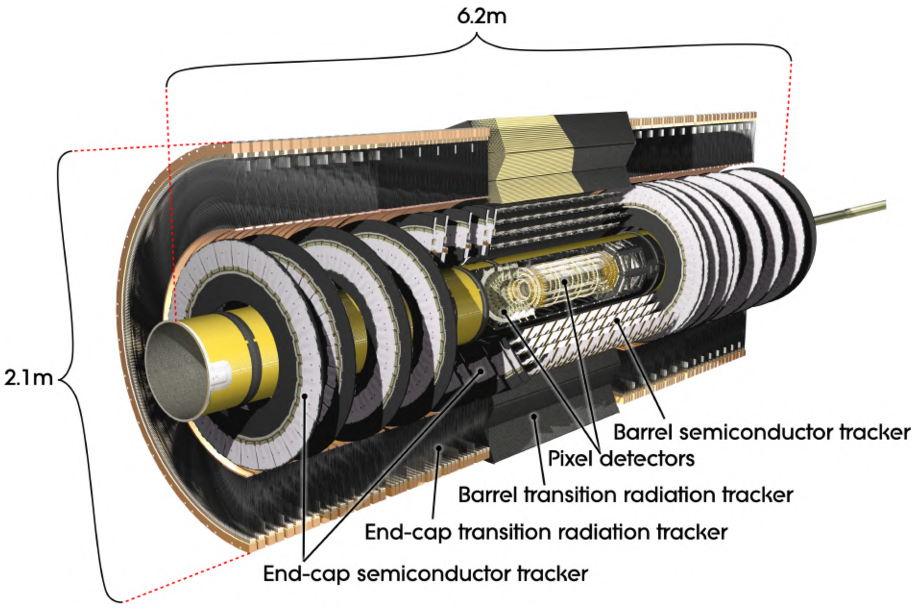


Figure 3.5: Cut-away view of ATLAS Inner Detector [61].

- Insertable B-Layer (IBL): the innermost layer, added in 2014, features radiation-hard sensors (50×250 μm) for high-luminosity conditions and improves vertex resolution [51].
- Pixel Detector: first detection point of the ATLAS experiment. It is made of 4 layers of silicon pixels, and it is characterized by a high granularity for accurate vertex and track reconstruction.
- Semiconductor Tracker (SCT): located around the pixel detector, it consists of more than 6 millions micro-strips of silicon sensors
- Transition Radiation Traker (TRT): made of many layers of straw tubes filled with $Xe/CO_2/O_2$ gas mixture, each with a diameter of 4 μm, interleaved with transition radiation material. The TRT allows electron identification via transition radiation processes.

3.2.3 Calorimeters

As shown in Figure 3.7, the ATLAS experiment employs two cylindrically nested sampling calorimeters. The innermost is designed to measure the electromagnetic (EM) shower generated by electrons and photon interacting with the absorber material. The

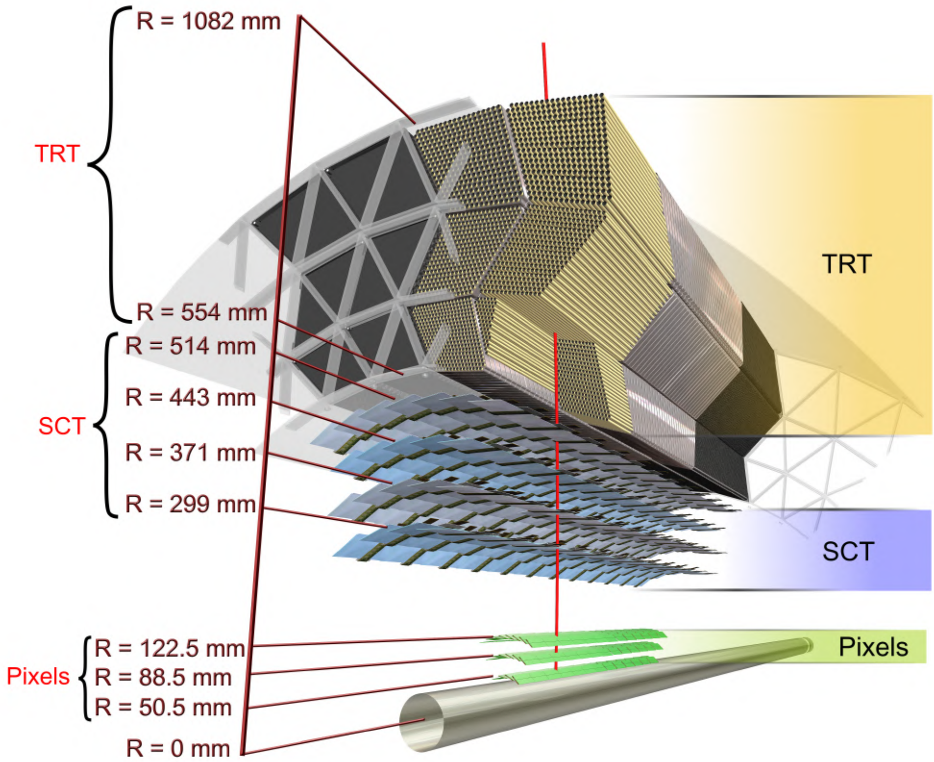


Figure 3.6: The structure of the ATLAS Inner Tracking Detector based on highly granular silicon pixels, silicon strips and straw tubes [61].

outer calorimeter measures hadronic shower, which results from hadron collisions with atomic nuclei. The wide range covered by both calorimeters corresponds to $|\eta| < 4.9$.

The depth of the calorimeter is a critical factor for its design, to ensure effective containment of both EM and hadronic showers. For EM calorimeters the relevant parameter is the radiation length (X_0), which corresponds to the average distance a high-energy electron cross through a material before its initial energy is reduced by a factor $1/e$ through *bremsstrahlung*. For hadronic calorimeters, the key parameter is the nuclear interaction length (λ), defined as the mean free path covered by the hadron before undergoing an inelastic interaction with a nucleus.

In the barrel region, the electromagnetic calorimeter has a total thickness exceeding 22 X_0, while in the end-cap regions it surpasses 24 X_0. The active calorimeter contributes approximately 9.7 λ in the barrel and 10 λ in the end-caps, providing sufficient depth to achieve accurate energy measurements for high-energy jets.

The electromagnetic calorimeter (ECAL) is a lead-liquid argon (LAr) detector that features accordion-shaped Kapton electrodes and lead absorber plates throughout its entire coverage. This accordion design ensures full azimuthal (ϕ) symmetry, eliminating any discontinuities in the ϕ direction [60]. The ECAL is segmented into a barrel section (covering the region $|\eta| < 1.475$) and two end-cap sections ($1.375 < |\eta| < 3.2$), each housed in its own cryostat, maintained at a temperature of $-184\,°\mathrm{C}$ to keep the argon in a liquid state.

The hadronic calorimeter (HCAL) surrounds the ECAL and is responsible for measuring the energy of hadrons, which are not fully absorbed by the inner layers. It consists of a central section, the Tile Calorimeter, covering $|\eta| < 1$, and two Hadronic End-cap Calorimeters (HEC) that extend coverage in the forward regions.

The Tile Calorimeter, positioned just outside the ECAL barrel, is a steel–scintillator sampling calorimeter covering up to $|\eta| < 1.7$ with a barrel and extended barrels segmented into 64 azimuthal modules and three radial layers, totaling up to 9.7 interaction lengths at $\eta = 0$. The HEC, located behind the end-cap ECAL and sharing its cryostat, spans $|\eta| = 1.5$ to 3.2 and uses copper plates interleaved with liquid argon in modular wheels. Finally, the Forward Calorimeter (FCal), integrated within the same cryostat, provides up to 10 λ of coverage using dense copper and tungsten modules with narrow LAr gaps, ensuring efficient energy measurement in the forward region and minimizing background radiation in the muon system [60].

3.2.4 Muons Spectometer

The outermost layer of the ATLAS detector is the Muon Spectrometer (MS), designed to identify muons, which traverse the calorimeters with minimal energy loss. It also measures the muon momentum by tracking the curvature of their paths in the presence of a magnetic field—typical values in the range of 0.5-2 T. For $|\eta| < 1.4$, the magnetic field is generated by a large barrel toroid, while in the region $1.6 < |\eta| < 2.7$, the muon trajectories are bend by smaller end-cap magnets [63].

The layout of the MS is presented in Figure 3.8. This configuration corresponds to the detector setup used during Run 2 of LHC operations.

The MS is composed of over 4,000 individual muon chambers, divided into separate system: trigger and tracking chambers.

The trigger chambers include Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs), covering a range of $|\eta| < 2.4$. Both chamber types deliver prompt signals, with nanosecond-level time resolution, enabling Level-1 triggering and precise bunch crossing determination.

An RPC is a gaseous detector with a simple structure. It consists of a gas mixture containing $C_2H_2F_4$ enclosed between two bakelite plates separated by polycarbonate spacers. When a muon passes through an RPC, ionization electrons undergo avalanche

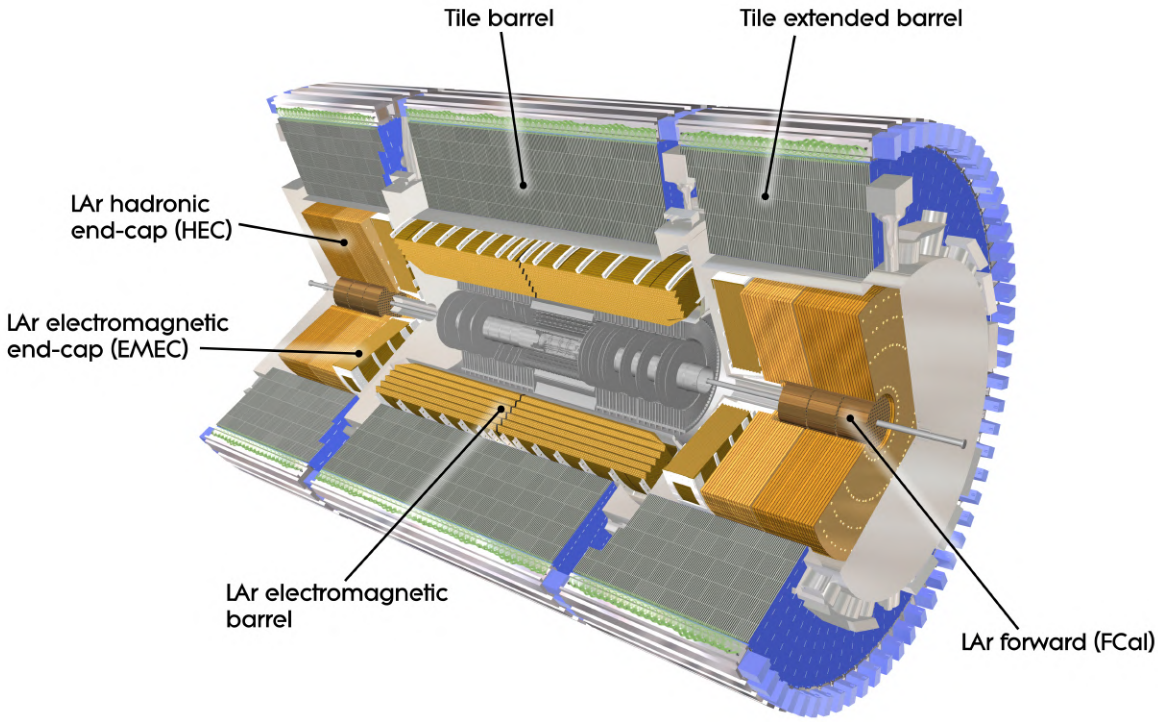


Figure 3.7: Cut-away view of ATLAS Calorimeter [62].

multiplication, producing a signal that is read out by orthogonal strips to measure the η and ϕ coordinates. The spatial resolution is approximately 1 cm, while the time resolution is about 1 ns.

TGCs are multi-wire proportional chambers operating near saturation with a CO_2–C_5H_2 gas mixture, and are used to measure the azimuthal coordinate (ϕ).

The tracking chambers are responsible for reconstructing the muon trajectory and are divided into Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs). The system is designed to ensure full coverage, such that each muon passes through at least three chambers.

MDTs, which operate in the region $|\eta| < 2$, are filled with a gas mixture that ionizes when traversed by a muon. By measuring the drift time, the trajectory and momentum of the muon can be reconstructed. In the region $2.0 < |\eta| < 2.7$, CSC chambers are used, offering higher granularity in the innermost layers to handle the higher event rates. CSCs provide a spatial resolution of 40 μm in the ϕ direction and 5 mm in η, with a time resolution of 7 ns [63].

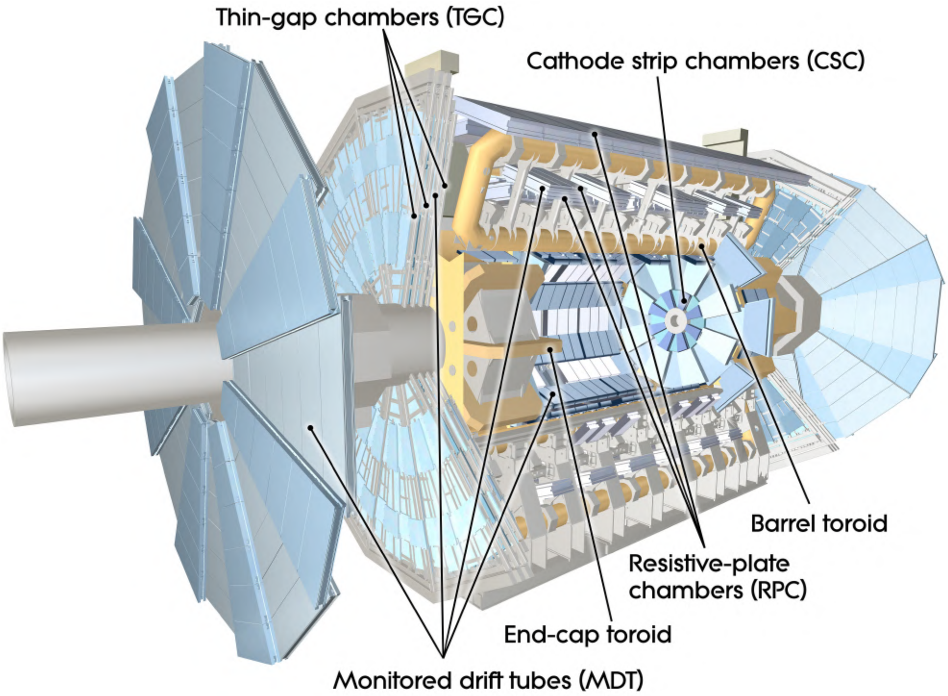


Figure 3.8: Cut-away view of ATLAS Muon Spectometer and main components [64].

3.2.5 Other Detectors

In addition to its main components, ATLAS includes four smaller detectors positioned at high rapidity within the ATLAS coordinate system. The primary function of the first two is the measurement of luminosity:

- LUCID (*LUminosity measurement using a Cherenkov Integrating Detector*), is situated at approximately ±17 meters from the interaction point. It is designed for high-precision luminosity measurements, both in real time and offline for physics analyses. LUCID is able of providing per-bunch absolute luminosity across various LHC beam conditions—including high-luminosity *pp* collisions, heavy-ion runs, and special low-luminosity configurations. The upgraded version, LUCID-2, introduced in Run 2 in 2014 and used in Run 3 as well, features improved radiation tolerance and built-in redundancy, allowing stable and accurate performance even at peak luminosities. It works alongside other independent luminosity monitoring systems (e.g. ID-, and LAr-based methods) to ensure cross-checks and calibration. Additionally, prototype detectors installed for Run 3 support both current measurements and research and development for the HL-LHC upgrade, LUCID-3 [51].

- ALFA (*Absolute Luminosity For ATLAS*) is positioned at ±240 meters from the collision point. It consists of scintillating fiber structures housed within Roman pots—special devices designed to be placed as close as possible to the beam in order to detect particles scattered at very small angles. Initially used during dedicated runs in Run 2 for elastic and diffractive studies, ALFA has been upgraded for operation in Run 3. Parts of the ALFA readout electronics were replaced due to aging and radiation damage. To mitigate further degradation in Run 3, iron shielding was installed, halving the radiation load and extending detector longevity [51].

- ZDC (*Zero-Degree Calorimeter*) is located at ±140 meters from the interaction point, beyond the point where the common beam pipe splits again into two separate vacuum tubes. This detector plays a key role in determining the centrality of heavy-ion collisions and is composed of alternating layers of quartz bars and tungsten plates. In the ATLAS heavy-ion program, the ZDC measures the energy of spectator neutrons—those not involved in the collision—providing crucial information on the impact parameter and nuclear breakup [51].

- AFP (*ATLAS Forward Proton*) is a dedicated detector system positioned symmetrically at ±205 m and ±217 m from the ATLAS interaction point, installed inside movable beam pipe sections called Roman pots. The detector enables the measurement of protons that remain intact after a collision and are scattered at very small angles relative to the beam-line. AFP consists of high-resolution silicon pixel trackers and time-of-flight detectors, which allow precise reconstruction of the proton's momentum and arrival time. This detector plays a crucial role in studies of central exclusive production and diffractive processes, enhancing sensitivity to rare interactions by tagging forward-scattered protons and suppressing background from non-diffractive events [65, 66].

3.2.6 Trigger and Data Acquisition Systems

The ATLAS trigger system allows online processing, selection, and recording of events with features of interest for subsequent offline physics analysis. The data selection process through the trigger system is divided into two successive stages: Level 1 (L1) and High-Level (HL). The triggered data are then driven by the Data Acquistion (DAQ) system from the subdetector electronics to offline processing [67].

First Level Selection (L1)

The Level-1 trigger is hardware-based and performs an initial selection based on data obtained from the calorimeters and the muon spectrometer. The L1 calorimeter (L1Calo) trigger identifies electrons, photons, and τ-leptons above a chosen threshold, select jet candidates, and calculates the missing transverse energy; the L1 muon (L1Muon) trigger detects the deviation of muons from the collision point.

The L1 trigger can retain events based on certain physical quantities, such as the total energy in the calorimeter, the number of objects above a fixed threshold (e.g., a certain muon p_T), or by applying topological requirements (e.g., invariant masses). Additionally, L1 trigger identifies Regions of Interest (RoIs) in η and ϕ, which will be further analyzed by the second-level trigger.

High-Level Selection (HL)

The second trigger stage, the HLT, is software-based and runs on a dedicated computing farm of about 40,000 Processing Units (PUs). Fast algorithms first reject uninteresting events in less than few hundred milliseconds, followed by more refined ones, similar to offline reconstruction, for final selection. Each step processes data within RoIs and ends with a hypothesis check based on extracted features. The farm is regularly upgraded to boost performance.

Chapter 4

Overview of the Analysis Strategy

This chapter provides a general overview of the analysis conducted to study hadronically decaying, boosted vector bosons (W/Z) produced in association with jets. The goal is to identify a resonant structure in the large-R jet mass spectrum consistent with the presence of boosted electroweak bosons, appearing as a peak on top of a smoothly falling background dominated by QCD multi-jet production.

To give a visual example of the expected signature, Figure 4.1 show the jet mass distribution from a previous ATLAS analysis performed at $\sqrt{s} = 7$ TeV [68], illustrating a resonant bump in the jet mass spectrum due to hadronic W/Z decays.

To highlight the composition of the background, it is important to note the stark difference in production cross sections: QCD multi-jet processes dominate, with cross sections several orders of magnitude larger than those of the electroweak V+jets signal, as visible in Figure 4.2. This disparity underscores the need for efficient background suppression techniques in analyses targeting hadronically decaying vector bosons [69].

4.1 Data Sample from MC Simulation

The study was based entirely on MC simulated dijet events generated with Pythia8 [70], corresponding to proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. These sample represent the full Run 2 configuration of the ATLAS detector, ensuring a detailed and realistic modeling of the detector response.

The event selection was defined by requiring at least one small-R jet with $p_{\mathrm{T}} >$ 25 GeV to identify the recoiling jet opposite the boosted large-R jet. Events with identified electrons or muons were vetoed. Additionally, at least one large-R jet with $p_{\mathrm{T}} > 500$ GeV was selected, a threshold chosen to ensure the events lie within the region where the lowest unprescaled jet trigger is fully efficient, minimizing trigger-related biases. The large-R jet mass was required to lie between 50 and 400 GeV. The mass selection $m > 50\,\mathrm{GeV}$ was applied because MC calibrations are unreliable below this

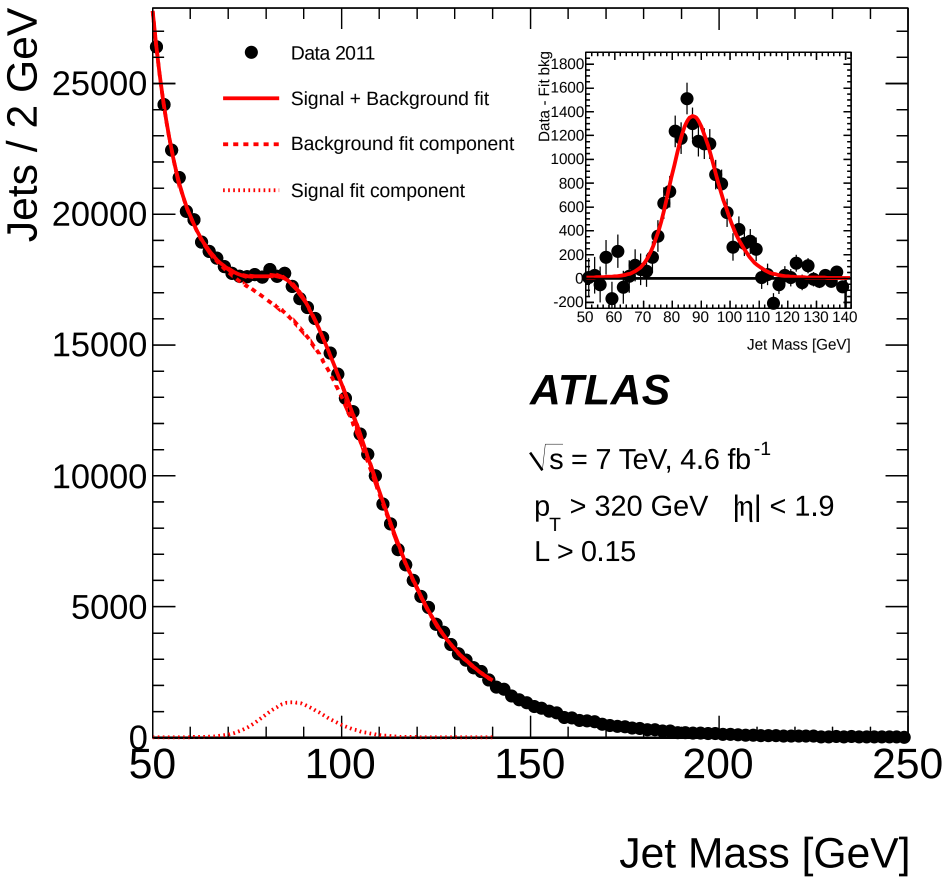


Figure 4.1: Jet mass distribution of selected W/Z jets overlaid with the fit result [68]. The background component (dashed), signal component (dotted), and total fit (solid) are displayed. The inset shows the difference between the data and the fitted background.

threshold due to significant detector effects and simulation uncertainties, ensuring accurate modeling and robust jet mass measurements.

In general, large-R jets were required to have a minimum transverse momentum of 200 GeV and a pseudorapidity $|\eta| < 2$ to ensure the boosted boson decay products were fully contained within the jet radius and well within the detector acceptance for reliable reconstruction.

The signal distribution was produced with Sherpa 2.2.14 [71], comprising a total of 25,000 events.

The histograms, based on the invariant mass distribution of the reconstructed boosted W/Z jet candidates, used in this study were generated using FastFrames [72], a lightweight framework designed for efficient ROOT ntuple processing. The framework

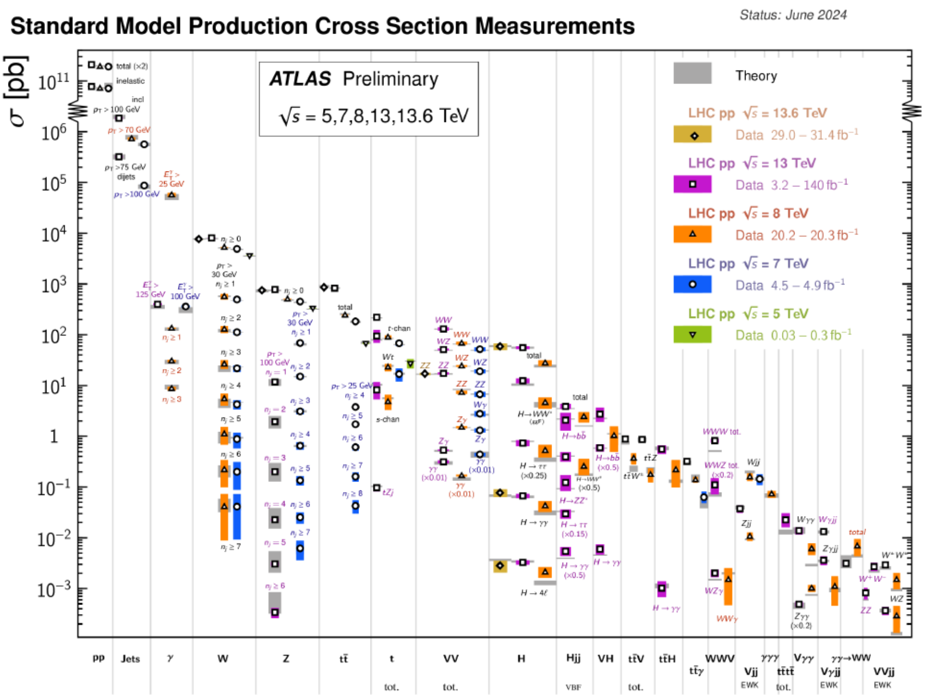


Figure 4.2: A summary of ATLAS Standard Model total and fiducial cross-section measurements in pp collisions at center-of-mass energies ranging from 5 TeV to 13 TeV [69].

applied the selection criteria described above to extract the relevant event samples.

As described in Section 3.1.2, pile-up arises in the detector due to multiple simultaneous proton–proton interactions occurring within a single bunch crossing. To ensure accurate modeling, event weights were applied to account for several corrections. First, the MC normalization weight is given by:

$$\text{Event weight} = \frac{\sigma \cdot \mathcal{L}}{N_{\text{gen}}} \tag{4.1}$$

where σ is the production cross section of the process, $\mathcal{L}$ is the integrated luminosity of the dataset, and N_{gen} is the total number of generated MC events [73]. Additionally, beam spot corrections were applied to compensate for any displacement or width differences between the beam spot in simulation and data. Pile-up reweighting was also performed to model realistic pile-up conditions based on expected distributions. Finally, Jet Vertex Tagger (JVT) efficiency scale factors were used to correct for differences in pile-up jet suppression performance between data and simulation efficiency scale factors [74, 75, 76].

In Figure 4.3, the background distribution for the leading[1] large-R jets in the selected

[1] The leading jet is defined as the jet with the highest transverse momentum p_T among those with

mass range from 50 to 400 GeV is shown, with the highlighted region indicating the chosen SR window, which is the region where a signal is expected to appear.

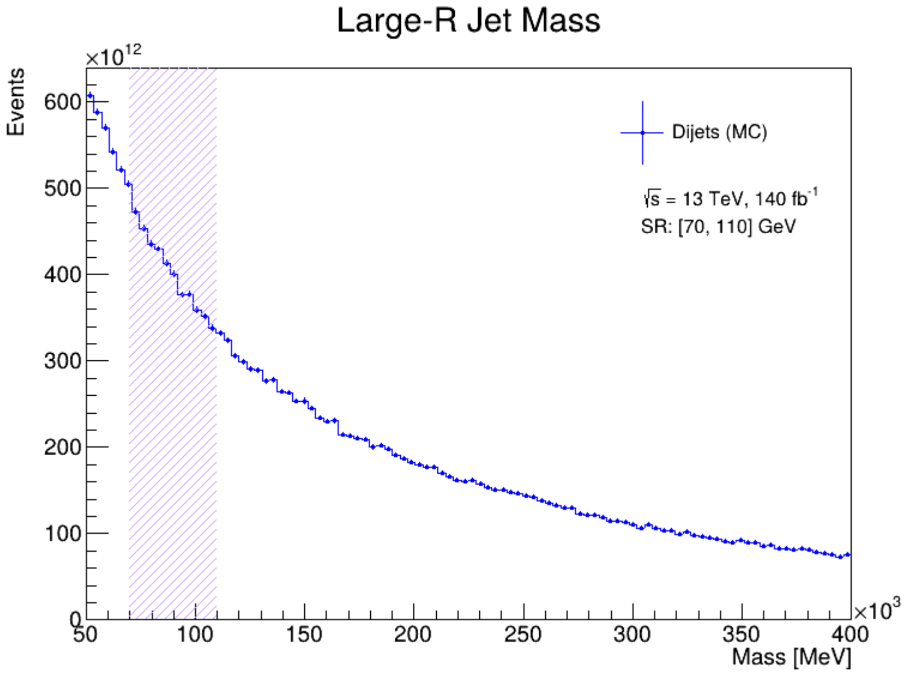


Figure 4.3: Background distribution with the SR [70, 110] GeV highlighted for the leading jets.

4.2 Signal Selection and Mass-Decorrelated Tagging

To enhance the identification of hadronically decaying vector bosons, an ANN classifier was applied in this analysis. The ANN is a machine-learning-based tagger that assigns a score to each jet, indicating how likely it is to originate from a W or Z boson decay. The tagger exploits a deep neural network architecture combining high-level jet substructure variables—such as N-subjettiness ratios, jet mass, energy correlation functions—and lower-level information like constituent kinematics, providing a rich representation of the jet's internal structure [33].

The classifier was operated at a working point corresponding to 50% signal efficiency, meaning that half of the true signal events were retained after the selection. This threshold was chosen as a compromise between suppressing the overwhelming QCD background

p_T larger than 500 GeV.

and retaining a statistically meaningful signal, thereby improving the overall signal-to-background discrimination. Importantly, this operating point results in a background rejection factor of approximately 10, reducing the QCD contamination by about 90%, which significantly enhances the purity of the selected jet sample without excessively compromising signal yield (see Figure 6 in [33]).

The classifier was trained to be decorrelated from the jet mass in order to avoid sculpting the background distribution into a shape that could mimic a resonant signal peak. While this mass decorrelation significantly reduces the risk of introducing bias, some residual dependence on the jet mass remains, potentially causing subtle distortions in the background shape after applying the tagger [77].

Chapter 5

Background Estimation

Among the possible outcomes of proton-proton collisions at the LHC are events containing two jets, known as dijet events.

As described in Section 2.4.1, hadronically decaying boosted W and Z bosons produce large-R jets, characterized by collimated particle showers resulting from their high transverse momentum. In such boosted topologies, the decay products are often merged into a single jet. Events containing a boosted vector boson often feature a second jet recoiling against it, balancing the transverse momentum.

In dijet analyses, the dominant background arises from standard QCD processes, which can conceal potential signals of interest. This background generally exhibits a smooth, rapidly falling behavior as a function of the invariant mass, whereas signals are expected to manifest as localized excesses or "bumps". Precise background estimation in the vicinity of the SR is therefore essential for identifying such anomalies.

This chapter focuses on background estimation using a sideband fitting technique, which is widely used in dijet resonance searches. It involves modeling the background using data from regions adjacent to—but explicitly excluding—the SR. In contrast, the Control Region (CR) is defined as a region with a negligible signal contribution and dominated by background processes. In scenarios where the signal manifests as a resonance peak (e.g., in invariant mass distributions), the SR is typically defined as a window around the resonance mass, while the CRs—often referred to as sidebands—are defined on either side of this mass window.

By fitting these background-dominated regions, it is possible to derive analytical or empirical model that accurately describes the background shape and can be reliably extrapolated into the SR.

5.1 Evaluation of Fit Performance

To evaluate how well a fit function describes the fitted distribution, the `TH1::Chisquare()` method provided by the ROOT data analysis framework was used [78]. This method computes the classical chi-squared statistic:

$$\chi^2 = \sum_i \frac{(y_i - f(x_i))^2}{\sigma_i^2},$$

where y_i is the observed value in bin i, $f(x_i)$ is the value of the fitted function at the center of bin i, and σ_i is the statistical uncertainty associated with y_i. The reduced chi-squared value, defined as χ^2/ndf, where `ndf` is the number of degrees of freedom, was computed to compare the goodness-of-fit across different functional forms and fitting ranges.

To visually complement the statistical evaluation, ratio plots beneath each histogram were included, showing the bin-by-bin ratio of the observed distribution to the fit function evaluated at the bin center. These plots offer a direct view of systematic deviations between the data and the model, highlighting trends or biases not immediately evident from the chi-squared value alone. A ratio close to 1 indicates good agreement, whereas consistent departures from 1 may suggest local mismodeling or inadequacies in the functional form.

5.2 Evaluation of Background Fit Models

As introduced in Section 5, the analysis involved fitting background distributions using various models and performing sideband fitting techniques.

The mass values within the SR were set to zero, as shown in Figure 5.1. Then a sideband fit was performed; each function was tested on the maximum range $[50 - 375]$ GeV, as well as on progressively narrower ranges, and the corresponding reduced chi-squared values (χ^2/ndf) were computed both globally and specifically in the signal region, defined as $[70 - 110]$ GeV.

Functions with 4 to 7 free parameters were explored, including sums of exponentials, polynomials, and hybrid models combining both functional forms. While increasing the number of free parameters improves the flexibility of the model in describing the background shape, it also introduces a greater risk of over-fitting, as a result of which the fit begins to capture statistical fluctuations rather than the underlying physical distribution.

Since the goal was to extrapolate the background distribution into the signal region, the high-mass tail was not considered essential for this study. Moreover, significant fluctuations in the ratio plot at high mass values suggest that the fit does not adequately describe the distribution in this region, as visible in Figure 5.2d.

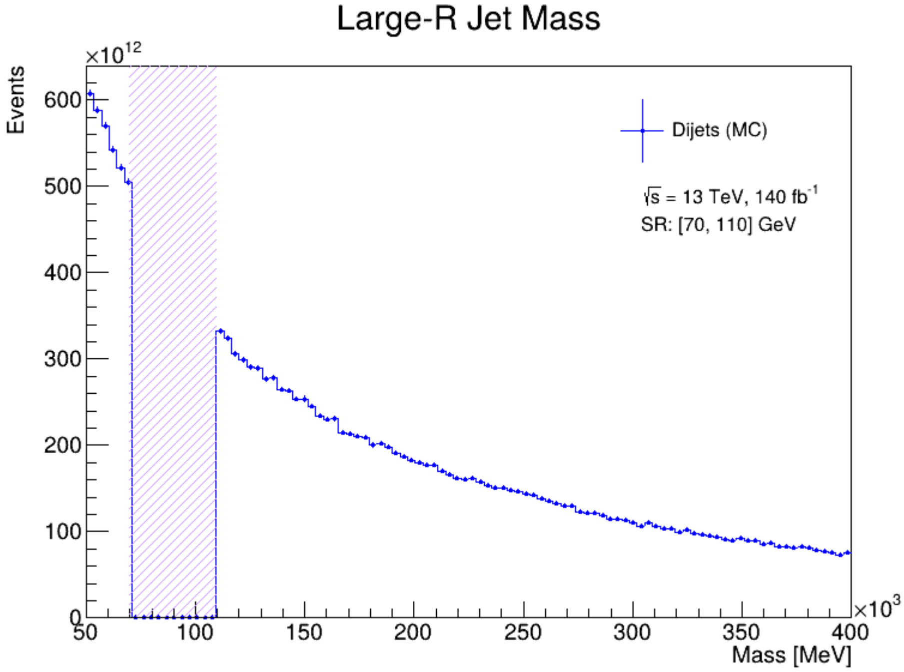
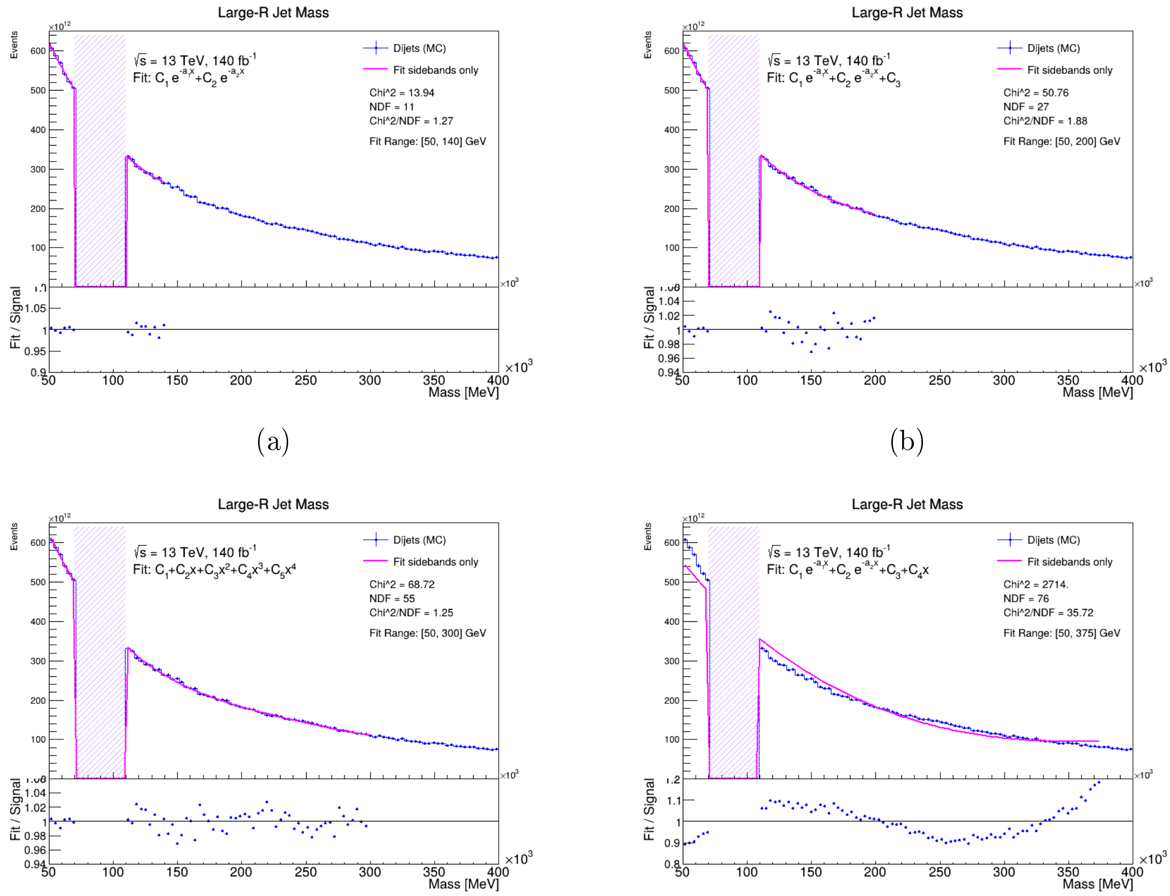


Figure 5.1: Background distribution with the SR $[70, 110]$ GeV blinded.

Some examples of the sideband fit results are presented in Figure 5.2, whereas in Figure 5.3, the fitted function is extrapolated into the SR using the parameters extracted from the fit.

In Table 5.1, the results of the fits obtained using various functional forms over different mass ranges are summarized. The results indicate that although many functions achieve a good overall fit quality, as reflected by $\chi^2/\mathrm{ndf} \approx 1$, their performance within the SR varies considerably. In particular, some functions show a substantial increase in χ^2/ndf within the SR, suggesting poor local agreement. These results were used to identify which models better capture the background shape while avoiding overfitting or poor interpolation in the SR.

In particular, the functional form consisting of the sum of two exponential terms (four-parameter case) provides a good description of the SR and was therefore selected for further analysis of the background distribution, as detailed in Section 6.1.



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

Figure 5.2: Examples of sideband-only fits to the large-R jet mass distribution using different functional forms and fit ranges. Each plot includes the fitted function, the corresponding χ^2/ndf, and a ratio panel showing the fit/data agreement outside the SR.

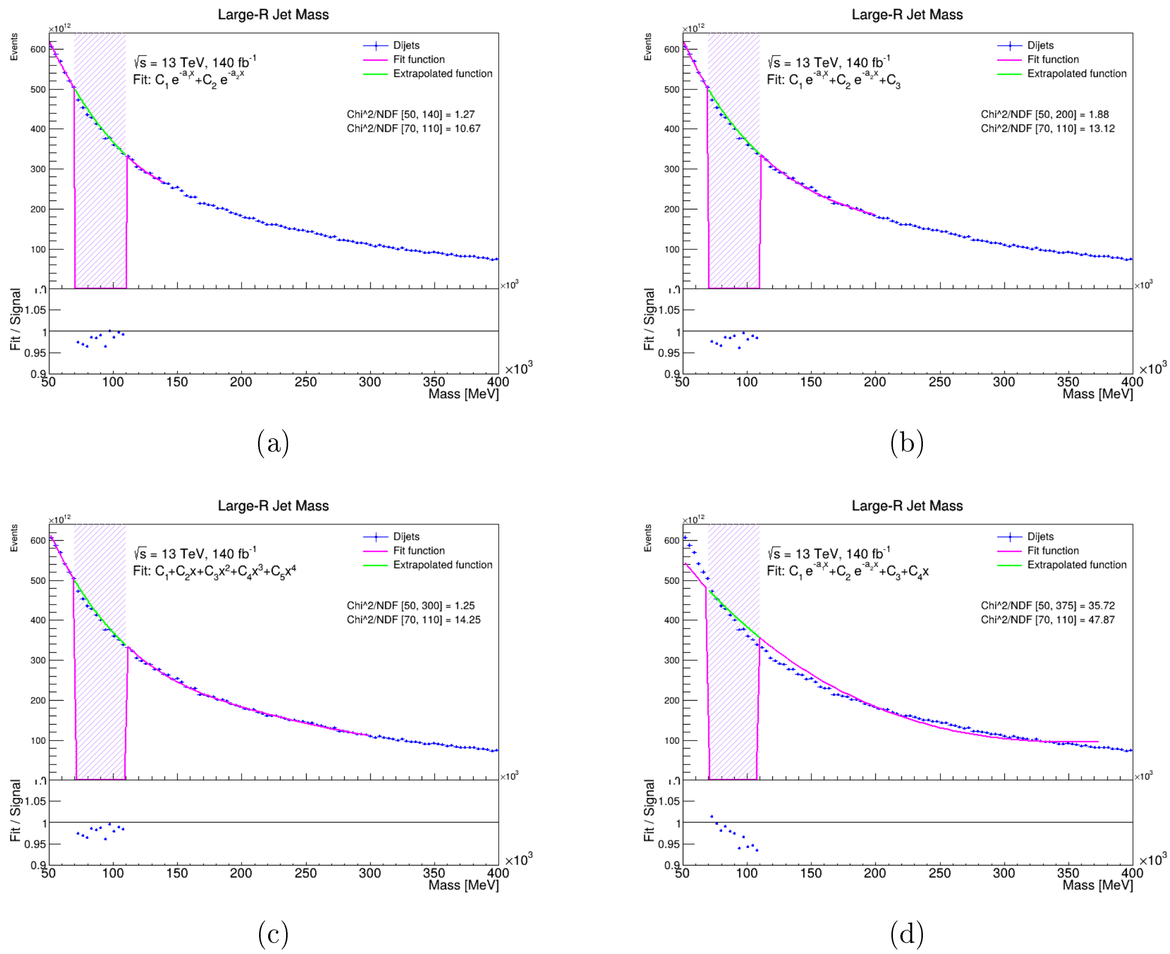


Figure 5.3: Extrapolation of the fitted background functions into the SR region (shown in green), based on the same distribution as in Figure 5.2. The corresponding χ^2/ndf values are computed using only the SR.

Fit function	**# parameters**	**Fit range (GeV)**	$\boldsymbol{\chi^2}$**/ndf**	$\boldsymbol{\chi^2}$**/ndf SR [70–110] GeV**
$C_1e^{-a_1x} + C_2e^{-a_2x}$	4	50–140	1.27	10.67
		50–200	1.63	5.97
		50–250	1.30	6.22
		50–300	1.27	4.17
		50–375	1.14	4.42
$C_1 + C_2x + C_3x^2 + C_4x^3$	4	50–140	1.27	10.07
		50–200	1.59	9.82
		50–250	2.04	21.92
		50–300	2.67	32.36
		50–375	9.80	48.05
$C_1e^{-a_1x} + C_2e^{-a_2x} + C_3$	5	50–140	1.67	7.90
		50–200	1.88	13.12
		50–250	2.10	21.11
		50–300	4.87	32.25
		50–375	9.53	42.87
$C_1 + C_2x + C_3x^2 + C_4x^3 + C_5x^4$	5	50–140	1.41	11.16
		50–200	1.60	8.12
		50–250	1.22	7.13
		50–300	1.25	14.25
		50–375	1.46	21.52
$C_1e^{-a_1x} + C_2e^{-a_2x} + C_3 + C_4x$	6	50–140	1.55	14.31
		50–200	1.82	12.24
		50–250	1.38	8.32
		50–300	1.22	7.28
		50–375	35.72	47.87
$C_1e^{-a_1x} + C_2e^{-a_2x} + C_3 + C_4x + C_5x^2$	7	50–140	1.75	17.68
		50–200	1.76	10.92
		50–250	1.37	8.96
		50–300	1.21	13.54
		50–375	1.08	10.60

Table 5.1: Fit results with χ^2/ndf values in different ranges for different functional forms.

5.3 Signal Distribution

After discussing the dominant QCD background, it is important to clarify how the signal was identified and categorized. The signal originates from hadronically decaying, boosted W and Z bosons, which are reconstructed as large-R jets. These jets are characterized by a distinct mass peak near the vector boson mass, allowing them to be distinguished from the smoothly falling QCD background.

To enhance signal-to-background discrimination, the large-R jets were further categorized using truth label selections. Events from the signal distribution were divided into three categories: (1) jets matched to vector bosons that are fully contained within the jet cone—meaning all decay products of the boson fall within the jet's radius (contained V), (2) jets matched to vector bosons whose decay products are only partially captured by the jet (non-contained V), and (3) jets not matched to any W/Z boson, which are assumed to originate from QCD background. These truth-level selections were applied during histogram generation using the FastFrames [72] framework.

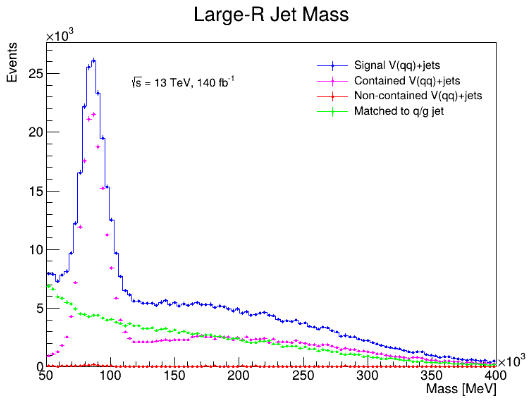
By isolating the vector boson components—especially the fully contained ones—it was possible to extract a region where the signal, comprising approximately 10^4 events, becomes more distinguishable against the dominant QCD background, which reaches up to 10^{14} events.

Figure 5.4 displays the individual contributions of the three categories to the total signal distribution for the leading p_T jet (the highest transverse momentum), as well as the sub-leading (the second highest) and sub-sub-leading jets (the third highest). While events are expected to contain primarily two jets from the boson decay and its recoil, additional jets such as the sub-sub-leading jets often arise from QCD radiation, including hard gluon emissions from the decay quarks or initial- and final-state radiation, resulting in extra quark- or gluon-initiated jets [79].

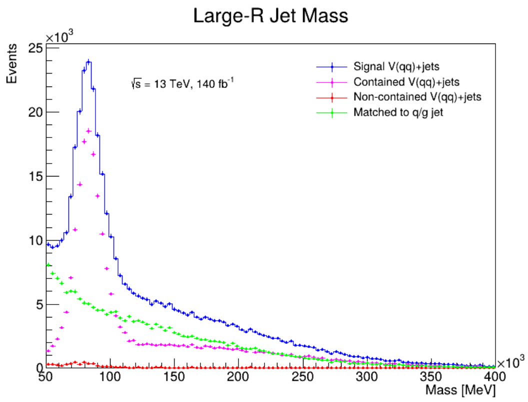
One notable feature is the presence of a long tail at high jet masses, particularly in the distributions of sub-leading and sub-sub-leading jets. This tail is believed to arise from the use of next-to-leading order (NLO) matrix element generation in the MC simulation, which introduced additional hard parton emissions and complex multi-jet topologies, enhancing the production of jets with higher masses and more intricate substructure.

In addiction, the overall signal distribution was analyzed with the aim of evaluating its contribution outside the initially defined SR, corresponding to the mass range of $[70, 110]$ GeV. Specifically, the signal histogram was integrated over some mass ranges beyond this window. This allowed to investigate the number of signal events falling beyond the selected mass window and better understand potential signal leakage outside the SR.

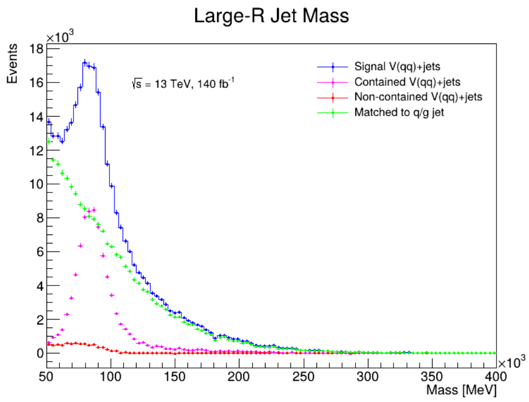
As shown in Table 5.2, approximately 40% of the signal events fall within the previously defined SR, which was applied exclusively to the leading large-R jet. However, it must be taken into account that the leading jet does not always correspond to the true vector boson decay product. In many signal events, the vector boson jet can be sub-



(a) Leading jets.



(b) Sub-leading jets.



(c) Sub-sub-leading jets.

Figure 5.4: "Contained" jets (pink) refer to those jets matched to vector bosons whose decay products lie entirely within the jet cone radius. In contrast, "non-contained" jets (red) jets are matched to vector bosons whose decay products are only partially captured by the jet. The green distribution represents jets not matched to W/Z bosons, thus considered background jets from QCD processes.

Event Region	**Total p_T**	**Leading p_T**
SR [70 - 110] GeV	0.444	0.403
SR [70 - 110] GeV (outside)	0.556	0.597
[70 - 120] GeV (outside)	0.517	0.563
[60 - 110] GeV (outside)	0.495	0.552
[60 - 120] GeV (outside)	0.456	0.518

Table 5.2: Summary of event fractions in different regions for both total and leading p_T signal distribution.

leading or even sub-sub-leading in p_T, particularly in the presence of a hard QCD recoil jet that dominates the event's kinematics. As a result, focusing solely on the leading jet causes the analysis to capture only a fraction of the total signal yield. Furthermore, for jet masses above 110 GeV, the dominant contribution to the total signal distribution originates from QCD-like jets, as illustrated in Figure 5.4a. To preserve a clean signal region and reduce contamination from background processes, the mass window was therefore not extended beyond this range.

Chapter 6

Background studies with ANN 50% Applied

The following chapter presents the results obtained from fitting the background-only and signal-plus-background distributions of the jet mass spectrum after application of the ANN tagger introduced in Section 4.2. These fits constitute a natural continuation of the analysis framework outlined in the previous chapter, which focused on identifying hadronically decaying, boosted vector bosons through their resonant signature in the large-R jet mass distribution. The significant disparity in production cross sections between signal and background necessitated the use of advanced background suppression techniques, such as the ANN tagger, to enhance sensitivity to the signal.

6.1 Fit Results for Background-Only and Signal-Plus-Background Models

In Figure 6.1, the signal-plus-background histogram is shown, after applying the ANN tagger to the leading jet.

To evaluate the modeling performance of the background and its behavior in the presence of a potential signal, several functional forms were tested. This section presents selected examples of the fits using three different models applied to both the background-only distribution and the signal-plus-background distribution.

The considered functional forms tested for fitting the mass distribution after applying the ANN tagger were:

1. A four-parameter model consisting of the sum of two exponential functions:

$$B(x) = C_1 e^{-a_1 x} + C_2 e^{-a_2 x} \tag{6.1}$$

This model was motivated by the results obtained in Section 5.2, where it demonstrated good descriptive power in the signal region.

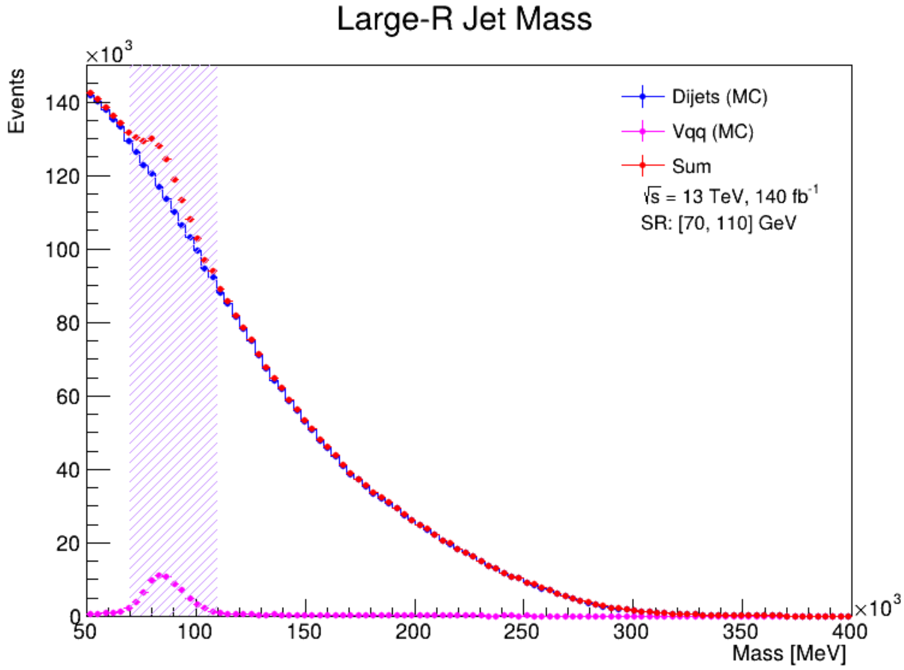


Figure 6.1: The signal distribution (pink) is added to the background distribution (blue) on a bin-by-bin basis, resulting in the total distribution shown in red, which corresponds to the leading jets.

2. A composite model including a signal-like component $S(x)$, with relative contribution f_E, in addition to the two exponentials:

$$B(x) = f_E \cdot S(x) + f_1 \cdot C_1 \cdot e^{a_1 x} + (1 - f_E - f_1) \cdot C_2 \cdot e^{a_2 x} \tag{6.2}$$

   where the sigmoid-like component $S(x)$ is defined as:

$$S(x) = \frac{\bar{x}}{\sqrt{1 + \bar{x}^2}} \quad \text{with} \quad \bar{x} = \frac{x - m_0}{\sigma_m}$$

   This model follows the approach described in Ref. [68], which successfully incorporated a signal-shaped term into the background fit.

3. A six-parameter model, which can be viewed as a simplified version of the second model, where the contributions from the two exponential components are controlled by a fraction parameter f_1:

$$B(x) = f_1 \cdot C_1 \cdot e^{a_1 x} + (1 - f_E - f_1) \cdot C_2 \cdot e^{a_2 x} \tag{6.3}$$

Two different upper bounds for the fit range were considered, corresponding to mass values of 140 and 170 GeV. This was done because, for mass values above 170 GeV, significant fluctuations were observed in the tail of the distribution.

After performing the fit on the background-only distribution using the CRs, the resulting background model was extrapolated into the signal region to estimate the background contribution under the signal peak. Subsequently, the same functional form(s) used to model the background in the sidebands were applied to fit the combined signal-plus-background distribution. This combined fit included an additional parameter μ_{sig} representing the signal strengh, such that the total fit function was expressed as:

$$f(x) = \mu_{sig} \cdot S(x) + B(x) \tag{6.4}$$

where $S(x)$ denotes the signal shape and $B(x)$ the background model.

Table 6.1 summarizes the performance of the three considered fit functions applied to the mass distribution after the ANN tagger selection. For each function, two fit ranges were explored, and the corresponding reduced chi-square (χ^2/ndf) and extracted signal strengh μ_{sig} are reported. The uncertainties on the fitted parameters were derived from the covariance matrix, which is provided by ROOT [78] and computed using the Minuit [80] minimization engine during the fit. The four-, six-, and eight-parameter models exhibit increasing flexibility, with the eight-parameter function explicitly modeling a potential signal-like contribution. As expected, the extended fit range generally leads to a higher χ^2/ndf, due to the increased difficulty of fitting a wider mass range.

Figure 6.2 shows the outcome of these fits: the left column displays the background-only fits, while the right column presents the corresponding signal-plus-background fits using the same functional form. A good agreement between the distribution and the models is observed outside the signal window, which justifies the extrapolation into the blinded region.

6.2 Spurious Signal Test

To evaluate potential biases in the fitting procedure, the full background-only distribution was fitted using the three signal-plus-background models described in Section 6.1. Since the data contains no true signal, this fit should ideally return a signal strength μ_{sig} consistent with zero. Any significant deviation from zero would indicate the presence of a spurious signal, revealing possible imperfections or overfitting in the chosen functional forms.

Figure 6.3 shows representative examples of the results obtained from this test, demonstrating consistency with the expected outcome.

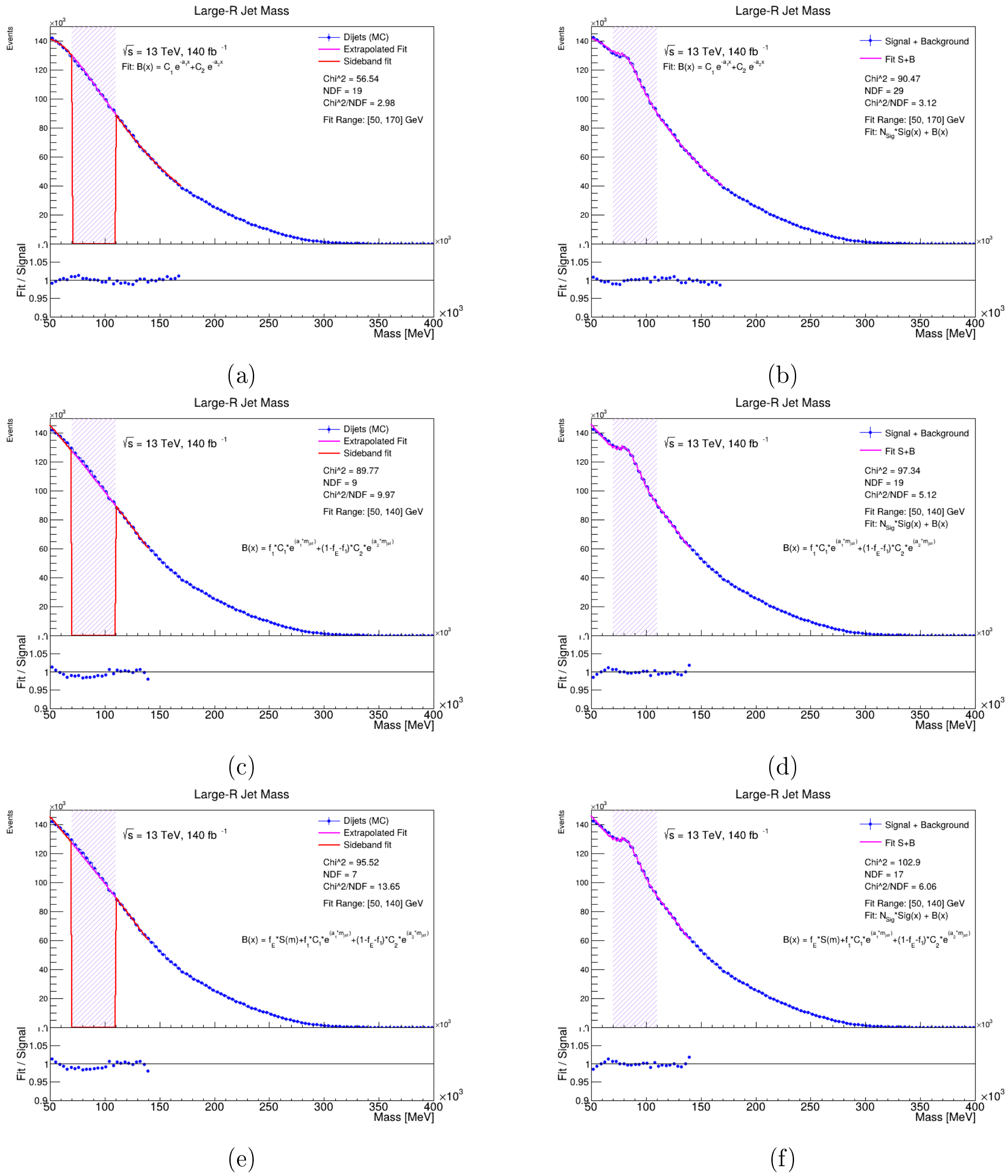
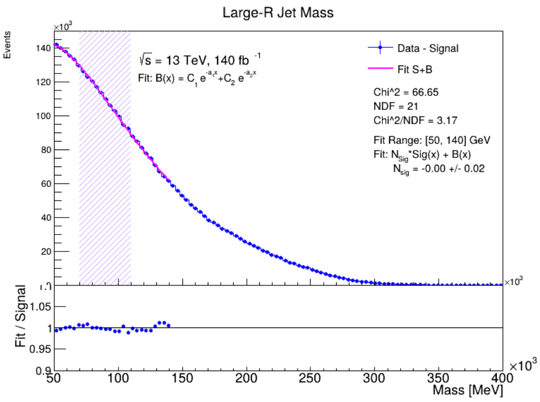


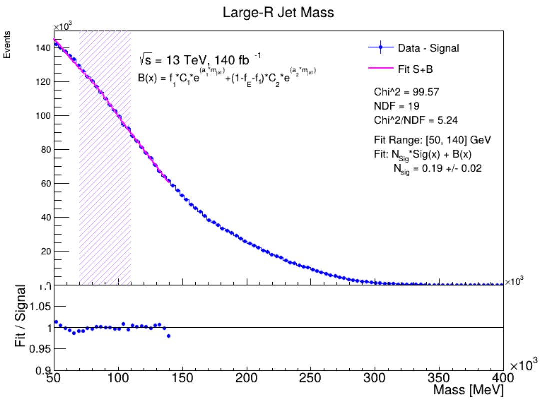
Figure 6.2: Overview of selected fit examples using three different models. The plots on the left show the background-only sideband fits with the extrapolation in the SR, while those on the right display the corresponding signal-plus-background fits using the same functional form.

Fit function $B(x)$	# parameters	Fit range (GeV)	χ^2/ndf	μ_{sig}
$C_1 e^{-a_1 x} + C_2 e^{-a_2 x}$	4	50–140	3.06	0.997 ± 0.023
		50–170	3.12	0.924 ± 0.023
$f_1 C_1 e^{a_1 x} + (1 - f_E - f_1) C_2 e^{a_2 x}$	6	50–140	5.12	1.188 ± 0.023
		50–170	11.67	1.403 ± 0.023
$f_E S(x) + f_1 C_1 e^{a_1 x} + (1 - f_E - f_1) C_2 e^{a_2 x}$	8	50–140	6.06	1.197 ± 0.029
		50–170	12.67	1.408 ± 0.023

Table 6.1: Fit results for different functional forms applied to the signal-plus-background distribution.



(a)



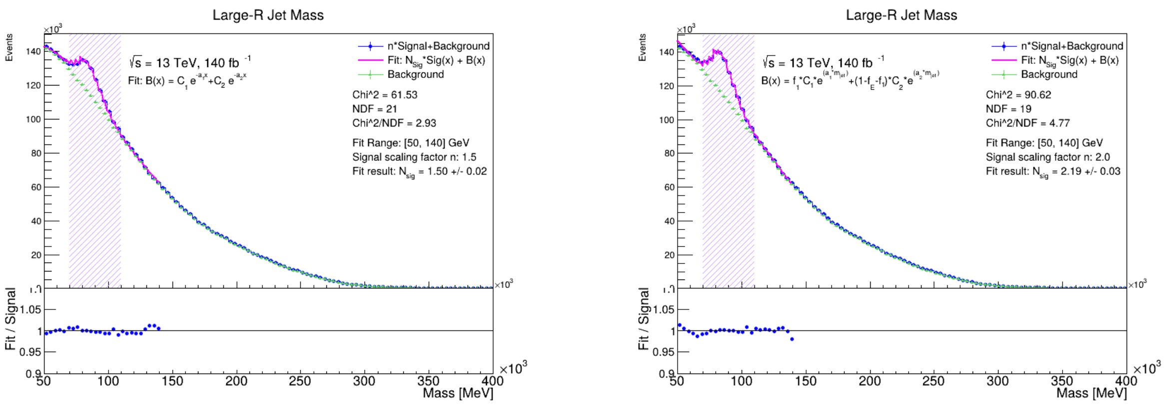
(b)

Figure 6.3: Examples of results from the spurious signal test. The quantity N_{Sig} shown in the plot represents the signal strength μ_{sig}.

6.3 Signal Injection Test

The signal injection test was performed by generating new distributions that combined the background distribution with scaled versions of the signal, using scaling factors ranging from 0.5 to 5 in increments of 0.5 times the original signal strength. Each modified distribution was refitted to extract the signal strength (μ_{sig}) using the three signal-plus-background models described in Section 6.1. With the extracted signal strength against the multiplicative scaling factor N, it was possible to verify the linearity and robustness of the fitting procedure in recovering the true signal strength.

Figure 6.4 presents selected examples of the signal injection test. In Figure 6.5, the results for all signal scaling factors are shown together, covering different combinations of fit parameters and maximum fit ranges. The model with 4 parameters corresponds to Equation 6.1, the one with 6 parameters to Equation 6.3, and the 8-parameter model to Equation 6.2. The maximum fit ranges considered were 140 GeV and 170 GeV. The plot shows that while all models exhibit approximately the same slope as the bisector, only those with fewer parameters and a lower maximum fit range produce results closer to the bisector line, indicating better accuracy in estimating the signal strength.



(a) (b)

Figure 6.4: Examples of results from the signal injection test. The quantity N_{Sig} shown in the plot represents the signal strength μ_{sig}.

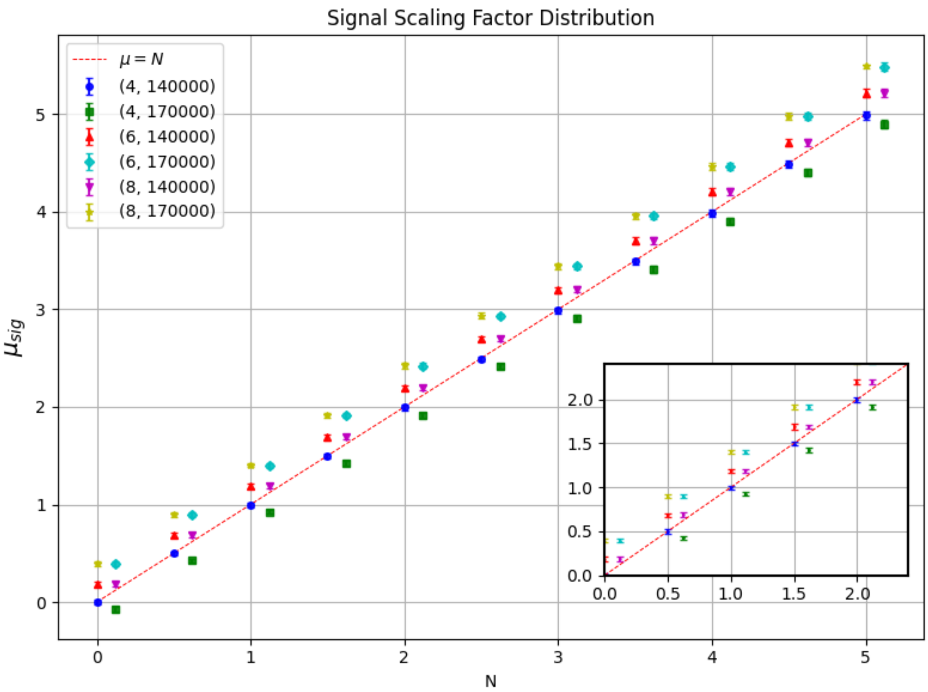


Figure 6.5: Distribution of the extracted signal strength (μ_{sig}) as a function of the signal scaling factor (N), shown for all fit functions and across different maximum fit ranges. The data points corresponding to the fit functions with colors cyan, green, and magenta have been horizontally shifted slightly for clarity. In the legend, each point indicates the number of fit parameters followed by the maximum fit range used.

Conclusion

This thesis has addressed the estimation of the dijet background in the context of hadronic V+jets cross section measurements, where the vector boson ($V = W, Z$) decays into quarks and is reconstructed as a large-radius jet. The accurate modeling of this background is crucial not only for testing predictions of perturbative QCD, but also for improving the sensitivity of searches for new physics phenomena where V+jets processes represent a major background. The complexity of the dijet background arises from the overwhelming contribution of QCD multijet processes, which produce smoothly falling distributions that can obscure potential signals manifesting as localized excesses.

A first approach implemented in this work was based on fitting smooth analytical functions to background-dominated CRs adjacent to the signal region. This sideband fitting technique exploits the smoothness of the QCD background invariant mass distribution to extrapolate the background shape into the signal region. In simpler cases, this method demonstrated robust performance and yielded good agreement with the distribution, validating the use of parametric functions for background modeling.

To further improve the discrimination between signal and background, this thesis explored the integration of machine learning techniques—specifically ANN. This allowed to perform a signal-plus-background fit on the full distribution, applying smooth functional forms.

Throughout the analysis, a uniform binning strategy was adopted for all histograms and distributions to ensure consistency across different regions of phase space. While this approach ensures consistency, it may not be optimal in all circumstances. In future analyses, adopting adaptive or non-uniform binning could help improve statistical stability in low-yield regions and enhance the sensitivity to signal features.

In conclusion, this thesis lays a solid foundation for the accurate estimation of dijet backgrounds in hadronic V+jets analyses, demonstrating the synergy between traditional fitting methods and modern machine learning techniques. The strategies and insights developed here contribute to the broader goal of precision measurements at the Large Hadron Collider.

Bibliography

[1] Konstantinos Theofilatos. *Vector boson production in association with jets and heavy flavor quarks at the LHC*. Tech. rep. Geneva: CERN, 2013. URL: https://cds.cern.ch/record/1547536.

[2] ATLAS Collaboration. “Study of (W/Z)H production and Higgs boson couplings using $H \to WW^*$ decays with the ATLAS detector”. In: *Journal of High Energy Physics* 2015.8 (2015), p. 137. DOI: 10.1007/JHEP08(2015)137.

[3] M. Krohn and C. Vernieri. *Search for new resonances decaying into boosted W, Z and H bosons at CMS*. 2017. arXiv: 1710.02217 [hep-ex].

[4] ATLAS Collaboration. “Searches for heavy diboson resonances in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2016.9 (Sept. 2016). ISSN: 1029-8479. DOI: 10.1007/jhep09(2016)173.

[5] ATLAS Collaboration. “Measurement of $W^{\pm}Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *The European Physical Journal C* 79.6 (June 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7027-6.

[6] Particle Data Group. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptac097.

[7] R. Fruhwirth and A. Strandlie. “Event Reconstruction”. In: *Pattern Recognition, Tracking and Vertex Reconstruction in Particle Detectors*. Cham: Springer International Publishing, 2021, pp. 23–31. ISBN: 978-3-030-65771-0. DOI: 10.1007/978-3-030-65771-0_2.

[8] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5031-2.

[9] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML]. URL: https://arxiv.org/abs/1406.2661.

[10] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020.

[11] Mark Thomson. *Modern Particle Physics*. Cambridge, UK: Cambridge University Press, 2013. ISBN: 9781107034266. URL: https://doi.org/10.1017/CBO9781139525367.

[12] Wikipedia contributors. *Standard Model of Elementary Particles — Wikipedia, The Free Encyclopedia*. Image under CC BY-SA 3.0. 2024. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (visited on 05/30/2025).

[13] Wolfgang Pauli. "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren". In: *Zeitschrift für Physik* 31 (1925), pp. 765–783. DOI: 10.1007/BF02980631.

[14] S. Braibant, G. Giacomelli, and M. Spurio. *Particelle ed interazioni fondamentali*. UNITEX. Milano: Springer, 2009.

[15] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[16] Kenneth G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445.

[17] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2013, pp. 603–607.

[18] National Institute of Standards and Technology. *2022 CODATA Value: reduced Planck constant*. The NIST Reference on Constants, Units, and Uncertainty. URL: https://physics.nist.gov/cgi-bin/cuu/Value?hbar.

[19] Chien-Shiung Wu. "The discovery of the parity violation in weak interactions and its recent developments". In: *Lect. Notes Phys.* 746 (2008), pp. 43–70. DOI: 10.1007/978-4-431-77056-5_4.

[20] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652.

[21] M. Battaglia et al. *The CKM Matrix and the Unitarity Triangle*. 2003. arXiv: hep-ph/0304132 [hep-ph]. URL: https://arxiv.org/abs/hep-ph/0304132.

[22] ATLAS Collaboration. *Inner Detector alignment development and performance in preparation for Run 3*. ATLAS PUB Note ATL-PHYS-PUB-2022-028. CERN, Geneva, Switzerland. June 2022. URL: https://cds.cern.ch/record/2811212/files/ATL-PHYS-PUB-2022-028.pdf.

[23] Anadijiban Das. “The Lorentz Transformation”. In: *The Special Theory of Relativity: A Mathematical Exposition*. New York, NY: Springer New York, 1993, pp. 48–71. ISBN: 978-1-4612-0893-8. DOI: 10.1007/978-1-4612-0893-8_3.

[24] S. Catani et al. “Longitudinally-invariant $k_{\perp}$-clustering algorithms for hadron-hadron collisions”. In: *Nucl. Phys. B* 406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M.

[25] M. Wobisch and T. Wengler. *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*. 1999. arXiv: hep-ph/9907280 [hep-ph].

[26] Yu.L Dokshitzer et al. “Better jet clustering algorithms”. In: *Journal of High Energy Physics* 1997.08 (Aug. 1997), pp. 001–001. ISSN: 1029-8479. DOI: 10.1088/1126-6708/1997/08/001.

[27] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti-k_t jet clustering algorithm”. In: *Journal of High Energy Physics* 04 (2008). Published 16 April 2008, p. 063. DOI: 10.1088/1126-6708/2008/04/063.

[28] ATLAS Collaboration. “Reconstruction of Jets”. ATLAS Open Data documentation. General Photo. URL: %7Bhttps://opendata.atlas.cern/docs/documentation/physic_objects/jets_reco%7D.

[29] Steven Schramm. *ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2291608.

[30] ATLAS Collaboration. “Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at sqrts = 8 TeV”. In: *Eur. Phys. J. C* (2016). URL: https://arxiv.org/pdf/1510.05821.

[31] ATLAS Collaboration. *Improving jet substructure performance in ATLAS using Track-CaloClusters*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-015. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2275636.

[32] ATLAS Collaboration. *Performance of small-radius $R = 0.4$ jets reconstructed from Unified Flow Objects in ATLAS*. ATL-PHYS-PUB-2022-038, ATLAS PUB Note. CERN, Geneva, Switzerland. Sept. 2022. URL: https://cds.cern.ch/record/2824558/files/ATL-PHYS-PUB-2022-038.pdf.

[33] ATLAS Collaboration. *Performance of W/Z taggers using UFO jets in ATLAS*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-029. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2777009.

[34] Andrew J. Larkoski et al. “Soft drop”. In: *Journal of High Energy Physics* 2014.5 (May 2014). ISSN: 1029-8479. DOI: 10.1007/jhep05(2014)146.

[35] ATLAS Collaboration. “Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC”. In: *Eur. Phys. J. C* 79.5 (2019), p. 375. DOI: 10.1140/epjc/s10052-019-6847-8. arXiv: 1808.07858.

[36] Jose Munoz, Ilyes Batatia, and Christoph Ortner. “Boost Invariant Polynomials for efficient Jet Tagging”. In: *Machine Learning: Science and Technology* 3 (Dec. 2022). DOI: 10.1088/2632-2153/aca9ca.

[37] CMS Collaboration. “Identification techniques for highly boosted W bosons that decay into hadrons”. In: *Journal of High Energy Physics* 2014.12 (Dec. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep12(2014)017.

[38] CERN. *The Large Hadron Collider*. Accessed: 2025-05-30. 2023. URL: https://home.cern/science/accelerators/large-hadron-collider.

[39] CERN. *Run 3: an opportunity to expand the LHC physics programme*. Accessed: 2025-05-30. June 2022. URL: https://home.cern/press/2022/run-3.

[40] CERN. *The Future Circular Collider*. Accessed: 2025-05-31. 2025. URL: https://home.cern/science/accelerators/future-circular-collider.

[41] Ewa Lopienska. “The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022”. In: (2022). General Photo. URL: https://cds.cern.ch/record/2800984.

[42] CERN. *Vacuum: Empty as Interstellar Space*. Accessed: 2025-05-31. 2023. URL: https://home.cern/science/engineering/vacuum-empty-interstellar-space.

[43] CERN. *Cryogenics: Low Temperatures, High Performance*. Accessed: 2025-05-31. 2023. URL: https://home.cern/science/engineering/cryogenics-low-temperatures-high-performance.

[44] ATLAS Collaboration. *About the ATLAS Experiment*. Accessed: 2025-05-31. 2023. URL: https://atlas.cern/about.

[45] CMS Collaboration. *The CMS Detector*. Accessed: 2025-05-31. 2025. URL: https://cms.cern/detector.

[46] ALICE Collaboration. *ALICE Physics Overview*. Accessed: 2025-05-31. 2023. URL: https://alice.cern/alice-physics.

[47] CERN. *LHCb Experiment*. Accessed: 2025-05-31. 2025. URL: https://home.cern/science/experiments/lhcb.

[48] Ana Lopes and Melissa Loyse Perrey. *LHC: The Guide*. English. CERN-Brochure-2021-004-Eng. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2809109.

[49] ATLAS Collaboration. “Measurement of the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{\rightarrow}4\ell$ cross-sections in pp collisions at sqrt(s) = 13.6 TeV with the ATLAS detector”. In: *The European Physical Journal C* 84.1 (Jan. 2024). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-12130-5.

[50] H. Burkhardt and P. Grafström. *Absolute Luminosity from Machine Parameters*. Technical Report. CERN, 2007. URL: https://cds.cern.ch/record/1056691/files/lhc-project-report-1019.pdf.

[51] ATLAS Collaboration. “The ATLAS experiment at the CERN Large Hadron Collider: a description of the detector configuration for Run 3”. In: *Journal of Instrumentation* 19.05 (May 2024), P05063. ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/05/p05063.

[52] ATLAS Collaboration. *ATLAS Public Luminosity Results from Run 1*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun1. Accessed: 2025-06-26. 2015.

[53] ATLAS Collaboration. *Public ATLAS Luminosity Results for Run 2*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (visited on 06/26/2025).

[54] CERN Commissioning Team. *LHC Long-Term Schedule*. Includes updates on Run 3, LS3, HL-LHC commissioning timeline. Sept. 2024. URL: https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm (visited on 06/26/2025).

[55] ATLAS Collaboration. *ATLAS Public Results: Offline Luminosity Analysis – Run 3*. Revision r37 (last update June 5, 2025). June 5, 2025. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3 (visited on 06/26/2025).

[56] ATLAS Collaboration. *Pile-up reweighting and configuration*. ATLAS Software Documentation, AnalysisSW Tutorial. Pileup analysis configuration via YAML. June 2025. URL: https://atlassoftwaredocs.web.cern.ch/analysis-software/AnalysisSWTutorial/cpalg_prw/.

[57] Joao Pequenao. “Computer generated image of the whole ATLAS detector”. General Photo. 2008. URL: https://cds.cern.ch/record/1095924.

[58] H.H.J. ten Kate. “The ATLAS superconducting magnet system at the Large Hadron Collider”. In: *Physica C: Superconductivity* 468.15 (2008). Proceedings of the 20th International Symposium on Superconductivity (ISS 2007), pp. 2137–2142. ISSN: 0921-4534. DOI: https://doi.org/10.1016/j.physc.2008.05.146.

[59] Ana Maria Rodriguez Vera and Joao Antunes Pequenao. “ATLAS Detector Magnet System”. General Photo. 2021. URL: https://cds.cern.ch/record/2770604.

[60] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[61] Joao Pequenao. "Computer generated image of the ATLAS inner detector". General Photo. 2008. URL: https://cds.cern.ch/record/1095926.

[62] Joao Pequenao. "Computer generated image of the ATLAS Muons subsystem". General Photo. 2008. URL: https://cds.cern.ch/record/1095929.

[63] ATLAS Collaboration. *ATLAS Muon Spectrometer: Technical Design Report*. CERN, 1997. URL: https://cds.cern.ch/record/331068.

[64] Joao Pequenao. "Computer generated image of the ATLAS calorimeter". General Photo. 2008. URL: https://cds.cern.ch/record/1095927.

[65] Adamczyk, L and Banaś, E and Brandt, A and Bruschi, M and Grinstein, S and Lange, J and Rijssenbeek, M and Sicho, P and Staszewski, R and Sykora, T and Trzebiński, M and Chwastowski, J and Korcyl, K. *Technical Design Report for the ATLAS Forward Proton Detector*. Tech. rep. 2015. URL: https://cds.cern.ch/record/2017378.

[66] Sebastian Grinstein. "The ATLAS Forward Proton Detector (AFP)". In: *Nuclear and Particle Physics Proceedings* 273-275 (Apr. 2016), pp. 1180–1184. DOI: 10.1016/j.nuclphysbps.2015.09.185.

[67] ATLAS collaboration. "Operation of the ATLAS trigger system in Run 2". In: *Journal of Instrumentation* 15.10 (Oct. 2020), P10004. DOI: 10.1088/1748-0221/15/10/P10004.

[68] ATLAS Collaboration. "Measurement of the cross-section of high transverse momentum vector bosons reconstructed as single jets and studies of jet substructure in pp collisions at sqrt (s) = 7 TeV with the ATLAS detector". In: *New Journal of Physics* 16.11 (Nov. 2014), p. 113013. ISSN: 1367-2630. DOI: 10.1088/1367-2630/16/11/113013.

[69] ATLAS Collaboration. *Expected performance of machine learning techniques for fully hadronic WZ and ZZ resonance searches at the HL-LHC*. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2024-011/. ATL-PHYS-PUB-2024-011. 2024.

[70] Christian Bierlich et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. 2022. arXiv: 2203.11601 [hep-ph]. URL: https://arxiv.org/abs/2203.11601.

[71] E. Bothmann et al. "Event Generation with Sherpa 2.2". In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[72] ATLAS Collaboration. *FastFrames documentation*. https://atlas-project-topreconstruction.web.cern.ch/fastframesdocumentation/latest/. 2024.

[73] Carsten Martin Magass. "Search for New Heavy Charged Gauge Bosons". PhD thesis. RWTH Aachen U., 2007. DOI: 10.2172/919581.

[74] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-018. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1700870.

[75] K. G. Tomiwa. "Performance of Jet Vertex Tagger in suppression of pileup jets and E_T^{miss} in ATLAS detector". In: *Journal of Physics: Conference Series* 802.1 (Jan. 2017), p. 012012. DOI: 10.1088/1742-6596/802/1/012012.

[76] ATLAS Collaboration. "Performance of pile-up mitigation techniques for jets in collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *European Physical Journal C* 76.10 (2016). Published 27 October 2016, p. 581. DOI: 10.1140/epjc/s10052-016-4395-z.

[77] ATLAS Collaboration. *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-014. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2630973.

[78] R. Brun and F. Rademakers. "ROOT — An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(97)00048-X.

[79] Michelangelo L. Mangano. "QCD and the physics of hadronic collisions". In: *Physics-Uspekhi* 53.2 (May 2010), p. 109. DOI: 10.3367/UFNe.0180.201002a.0113.

[80] F. James and M. Roos. "MINUIT: A System for Function Minimization and Analysis of the Parameter Errors and Correlations". In: *Comput. Phys. Commun.* 10 (1975), pp. 343–367. DOI: 10.1016/0010-4655(75)90039-9.