Università di Genova

Dipartimento di Matematica e Fisica
PhD Course in Physics and Nanoscience
XXXVIII cycle

Boosting Higgs Boson Physics at High Momenta and a New End-to-End Approach to Muon Tracking at the LHC

PhD Candidate
Lucrezia Rambelli

Supervisor
Dott. Andrea Coccaro
External Supervisor
Dott. Francesco Armando Di Bello

Contents

Abstract

In 2012, at the Large Hadron Collider, the largest particle accelerator built by humans, the Higgs boson was discovered by the ATLAS and CMS collaborations, being the last discovered piece of the Standard Model. With a measured mass of approximately 125 GeV and no spin, the Higgs boson has the unique feature of being the only known fundamental scalar particle. Its unique role originates from the structure of its scalar potential, whose characteristic symmetry-breaking shape leads to a non-zero vacuum expectation value, resulting in electroweak symmetry breaking and enabling mass generation for elementary particles while maintaining Lorentz invariance.

After the discovery, the ATLAS and CMS collaborations continued the analyses with the increasing datasets provided by the LHC entering into a precision era of Higgs-boson property measurements. Among its properties, the Higgs boson's couplings to other particles is being studied with high precision, since even small deviations from the Standard Model predictions could provide hints for new physics. The Higgs boson decay in a beauty quark pair, having a branching fraction of approximately 58%, has a dominant influence on the total Higgs width, and precise measurements of the Yukawa coupling of the Higgs boson to b quarks provide one of the most sensitive probes for looking at deviations from the Standard Model predictions.

In this context, a particularly suitable channel to study this coupling is the production of the Higgs boson in association with a vector boson VH (with $V = W, Z$ bosons), followed by its $H \to b\bar{b}$ decay. In fact, leptonic decay channels of the accompanying vector boson V provide clean experimental signatures which allow for a strong suppression of the otherwise overwhelming background processes.

As potential deviations from the Standard Model are enhanced at high energy scales, measurements in the boosted Higgs regime provide increased sensitivity to modifications of Higgs interactions, where higher-order effective operators introduce momentum-dependent corrections. In this regime, the cross section hierarchy changes with respect to the inclusive one where gluon-gluon fusion is nearly 90% of the total, and the contribution of VH production becomes increasingly significant while also offering clean leptonic signatures that improve the selection of $VH, H \to b\bar{b}$ events.

As a correct classification of the final state is fundamental for high precision measurements, over the years, several of the so-called flavour-tagging algorithms, i.e. algorithms able to identify and classify the flavour of the quark which produced the final state jets, were developed. Considering that in the boosted regime the two b-quarks in the final state become highly collimated, leading to two overlapping jet signatures, a dedicated flavour-tagging procedure is needed to accurately identify and classify the flavour of the resulting jets.

In fact, while at low energies the two final state jets are reconstructed and classified separately, in the boosted case the final state is reconstructed as a single jet that contains the products of both b-quark hadronizations.

The most recent algorithm developed by the ATLAS Collaboration is the GN2X tagger, a transformer-based model that relies exclusively on track-level information within the reconstructed large-radius jet to classify the flavour of the full reconstructed jet. Thanks to its superior performance compared to previous approaches, and even to state-of-the-art algorithms used in regimes where the two jets can be resolved separately, its use in analyses targeting $H \to b\bar{b}$

events can significantly enhance signal sensitivity, leading to more precise measurements of the Higgs coupling to b-quarks.

In this context, this thesis focuses on the identification of processes in which the Higgs boson decays into a pair of b-quarks in the boosted regime. During my PhD, my work has been dedicated first to the calibration of the GN2X tagger and subsequently to its integration into physics analyses targeting boosted $VH, H \to b\bar{b}$ final states.

Because the tagger is developed and tested on Monte Carlo samples where its behavior may differ from real data, it must be calibrated before use in physics analyses so that its performance in simulation matches that observed in actual data. Two calibration strategies were developed and implemented in this thesis for this preliminary study. The first strategy is purely based on Monte Carlo simulations, comparing tagger performance across different MC samples where systematic variations of input variables, driven by auxiliary measurements of tracking variables, are applied. This method proves particularly valuable for calibrating the tagger over processes like multijet backgrounds, where the simulated flavour fraction compositions suffer from large uncertainties.

The second calibration approach exploits the data collected by the ATLAS experiment in between 2015 and 2018, focusing on $Z \to b\bar{b}$+jets events as a high-statistics proxy for $H \to b\bar{b}$ processes. This choice is strategically motivated: $Z \to b\bar{b}$ is a color-singlet resonance with a mass close to that of the Higgs boson, making its b-jet kinematics very similar to the $H \to b\bar{b}$ signal, while offering a much higher available statistic. This data-driven approach allows direct comparison of tagger efficiency between simulation and data, yielding an independent set of scale factors that account for residual modeling uncertainties.

After obtaining these correction factors and their associated systematic uncertainties, I worked on integrating the GN2X tagger into the physics analysis targeting boosted decays of the Higgs boson in $VH, H \to b\bar{b}$ processes.

I developed the analysis framework in the boosted regime in order to incorporate the GN2X tagger into the signal event preselection, together with the corresponding correction factors derived from the calibration studies. A multivariate analysis was then performed to discriminate signal from background events using boosted decision trees (BDTs), whose output score distributions constitute the primary inputs to the final statistical fit and therefore play a central role in determining the analysis sensitivity.

While the full statistical inference was not performed within the scope of this work, the expected sensitivity was evaluated by comparing the BDT output score distributions obtained when including the GN2X tagger in the event preselection to those from previous ATLAS results. This comparison demonstrates a substantial improvement with respect to earlier analyses, with the enhanced sensitivity driven by the tagger's superior background rejection capabilities while maintaining high signal efficiency across the relevant kinematic phase space.

Motivated by the strong impact of the tagger at the event selection level, additional studies were carried out to investigate the potential benefit of incorporating GN2X-related information not only in the preselection but also directly as input variables to the BDT for each event. These studies were designed to assess whether a tighter integration of the tagger information within the multivariate framework could further enhance the analysis performance, providing indications of possible improvements beyond the current approach.

The final part of the thesis presents a research project that explores a differentiable end-to-end machine-learning approach to charged-particle track reconstruction, intended as a proof of concept demonstrating the potential of this paradigm. In particular, the proposed approach adopts differentiable programming techniques to enable the joint training of the entire reconstruction pipeline, allowing physics-motivated priors to be incorporated directly into the training process. This approach differs from conventional tracking strategies. In standard reconstruction pipelines, including those that incorporate machine-learning techniques, the tracking

problem is typically factorized into multiple sequential steps, such as hit selection, seeding, pattern recognition, and track fitting.

In contrast, the model that I developed formulates each stage of the reconstruction procedure in a differentiable way, allowing gradients to propagate through the successive steps of the workflow during training, enabling the different components of the model to share parameters and to be optimized jointly with respect to the global objective. As a result, the training process is aware of the role and impact of each step in the reconstruction chain, enabling a truly end-to-end optimization. The results show that this jointly optimized approach can outperform a more traditional factorized strategy, highlighting the advantages of treating track reconstruction as a fully differentiable learning problem.

The thesis is organized as follows. Chapter 1 introduces the theoretical foundations of the Standard Model, covering quantum electrodynamics, quantum chromodynamics, the electroweak sector, and the Brout-Englert-Higgs mechanism, as well as Higgs boson phenomenology at the LHC. Chapter 2 presents the Large Hadron Collider complex and the ATLAS detector, describing the sub-systems involved in tracking, calorimetry, and muon detection, which provided the data used in the analyses. In this Chapter also an explanation of the reconstruction of physics objects in ATLAS is given, with particular emphasis on jet reconstruction methodologies, since jets are the experimental signature of $b\bar{b}$ final states. Chapter 3 reviews jet flavour tagging strategies, introducing historical approaches and presenting GN2X, which is specifically optimized for boosted topologies where jets are highly collimated. Chapter 4 details the calibration of GN2X, including both Monte Carlo–based and data-driven approaches, and the derivation of scale factors necessary for incorporating the tagger into physics analyses. Chapter 5 describes the boosted $VH, H \to b\bar{b}$ analysis, explaining the importance of this channel, how the analysis is organized, and presenting how the calibrated GN2X tagger improves sensitivity and event categorization. Finally, Chapter 6 presents the differentiable end-to-end tracking framework, including model design, training, and performance evaluation, highlighting the integration of physics knowledge into machine learning for improved reconstruction performance.

Chapter 1

Theoretical Overview

The study of the smallest building blocks of nature has led to the development of a comprehensive theory known as the Standard Model (SM) of particle physics. This framework unifies our understanding of how matter and forces behave at the most fundamental level. It describes all known elementary particles and the interactions between them within a single mathematical structure. The theory emerged gradually through decades of theoretical and experimental discoveries; from the identification of the electron and the proton, to the development of quantum mechanics and the formulation of quantum field theory.

The Standard Model successfully combines three of the four fundamental forces of nature: the electromagnetic, weak, and strong interactions, without including gravity. These forces are mediated by particles called bosons, which act as carriers of the interactions between matter particles, known as fermions. The unification of these forces in a consistent theoretical framework represented one of the major achievements of twentieth-century physics.

The SM successfully predicted the existence of the top and bottom quarks, the W and Z bosons, and the tau neutrino. The final missing component of the theory, the Higgs boson, was first proposed in the 1960s and was eventually observed in 2012 at the Large Hadron Collider by the ATLAS and CMS experiments [1][2].

Since its discovery, extensive experimental efforts have focused on precise measurements of the Higgs boson mass and its interactions with other particles.

In this chapter, a general overview of the Standard Model is presented in Sections 1.1 and 1.2, while a more detailed discussion of the Higgs sector and its phenomenology is provided in Section 1.3.

1.1 The Standard Model of particle physics

The Standard Model is formulated within the framework of Quantum Field Theory (QFT). Within the QFT, every type of particle is associated with a quantum field that extends throughout space and time. What we observe as a particle corresponds to a localized excitation, or a "quantum," of its respective field. These fields are treated mathematically as operator-valued distributions acting on spacetime, allowing the probabilistic and wave-like nature of particles to be described in a unified theoretical framework.

At the heart of QFT lies the Lagrangian density, which encapsulates both the kinematic and dynamic properties of a field. The complete Lagrangian of the Standard Model can be separated into distinct components, each corresponding to a specific sector of the theory:

$$\mathscr{L}_{\text{SM}} = \mathscr{L}_{\text{EW}} + \mathscr{L}_{\text{QCD}} + \mathscr{L}_{\text{Higgs}} + \mathscr{L}_{\text{Yukawa}}. \tag{1.1}$$

$\mathscr{L}_{\text{SM}}$ is invariant under local gauge transformation of the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. Gauge

symmetries leave observable properties of the system unchanged when the corresponding gauge transformation are applied to the fields.

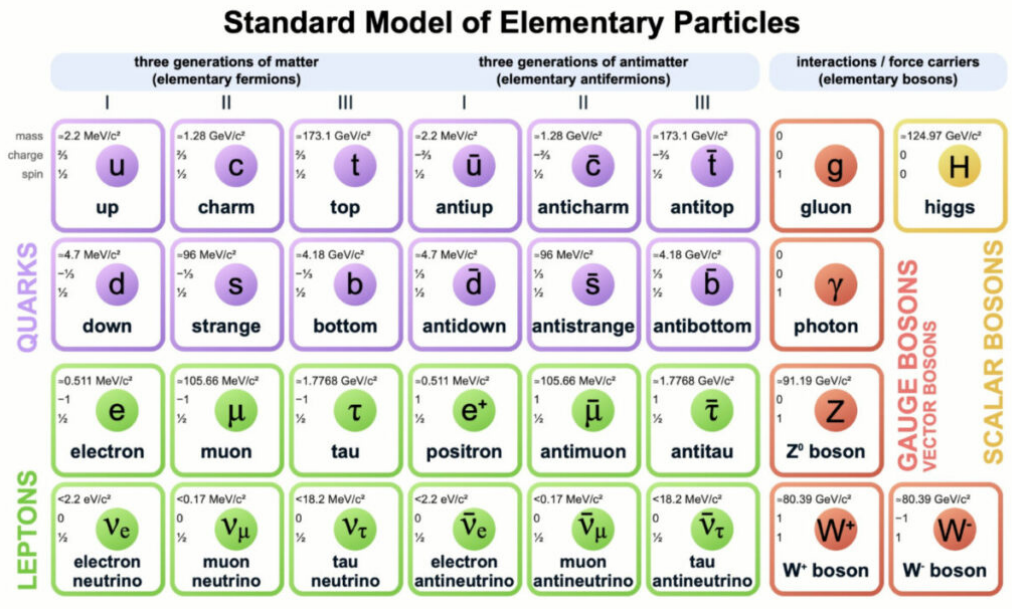


Figure 1.1: Particle content of the Standard Model of particle physics.

The particle content of the SM is depicted in Figure 1.1, and in the following paragraphs each term of the Equation 1.1 is described separately.

As the simplest and historically first experimentally verified sector of the Standard Model, Quantum Electrodynamics represents the Abelian gauge theory associated with the electromagnetic interaction and provides a natural starting point for discussing the individual components of the Standard Model Lagrangian.

1.1.1 Quantum Electrodynamics

Quantum electrodynamics (QED) is the relativistic quantum theory which describes the interaction between the photon (γ) and the charged electrically charged fermions.

Considering a Dirac spinor field $\psi = \psi(x)$ and its adjoint representation $\bar{\psi} = \psi^\dagger\gamma^0$, with $\psi^\dagger$ Hermitian conjugate of the ψ field, the Dirac Lagrangian density is

$$\mathscr{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi \ , \tag{1.2}$$

where ψ represent the fermionic spin-1/2 particle, for example an electron, μ is the Lorentz index ranging from 0 to 3, ∂_μ denotes the Lorentz-covariant derivative and γ^μ are the Dirac matrices.

The first term, $\bar{\psi}i\gamma^\mu\partial_\mu\psi$, represents the *kinetic term*, which governs the propagation of the fermion field through spacetime. The second term, $-m\bar{\psi}\psi$, is the *mass term*, coupling the left- and right-handed components of the field and giving rise to a particle mass m [1].

$\mathscr{L}_{\text{Dirac}}$ is invariant under U(1) *global* gauge transformation

$$\psi \to \psi' = e^{-ie\alpha}\psi \ , \tag{1.3}$$

[1] In the Standard Model, the left- and right-handed components of fermions transform differently under the $SU(2)_L \times U(1)_Y$ gauge symmetry. Because of this, the Dirac mass term is not invariant under electroweak gauge transformations, and explicit mass terms for fermions cannot be introduced directly without breaking the symmetry.

where e is a transformation parameter, but is not invariant under U(1) *local* gauge transformation, i.e. when the transformation depends on the spacetime coordinates $x = (x^0, x^1, x^2, t)$

$$\psi \to \psi' = e^{-ie\alpha(x)}\psi \ . \tag{1.4}$$

To ensure the local gauge invariance, a gauge-covariant derivative is introduced to substitute the ∂_μ term in the Equation 1.2 and is defined as

$$D_\mu = \partial_\mu + ieA_\mu \ . \tag{1.5}$$

Here, A_μ is a new added spin-1 field which transforms as $A_\mu \to A_\mu - \partial_\mu\alpha(x)$, and whose introduction results in a transformed interaction term which will cancel the asymmetric term in the *local* gauge transformation and a new interaction term between fermions and this new auxiliary field A_μ.

The kinetic term of the A_μ field, which can be identified with the electromagnetic one, is included by defining the electromagnetic tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, which is trivially invariant under the local gauge transformation introduced before. Including all this changes into the Equation 1.2, the electromagnetic kinetic term is incorporated into the QED Lagrangian term of the SM as

$$\mathscr{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi \tag{1.6}$$

.

The new added spin-1 field A_μ, introduced requiring the invariance under the local U(1) gauge transformation, is interpreted as the photon field, which interacts with charged matter. The quadratic mass term $A_\mu A^\mu$ related to it is not invariant, and therefore the field A_μ is required to be massless within the SM.

1.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) describes quarks, gluons, and their interactions. These particles carry colour charge, which comes in three types: red, green and blue. While QED was invariant under U(1) symmetry group transformation, the QCD Lagrangian is specified by requiring invariance under local SU(3) gauge transformations, making it a Yang-Mills theory in which the self-interaction of the gauge bosonig fields is required in addition.

The SU(3) group generators are given by $T_a = \lambda_a/2$, where λ_a are the eight Gell-Mann matrices. Each of the three up- and down-type of quark fields $q_k(x)$ of the SM transforms under the fundamental triplet representation of SU(3), in which each component of the triplet corresponds to the colour quantum number for red, green, and blue charges.

The gluon field strength tensors $G^a_{\mu\nu}$ are defined for each of the eight T_a generators, where each of them is defined as

$$G^a_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - g_s f^{abc} A^b_\mu A^c_\nu \tag{1.7}$$

with A^a_μ gluon fields, g_s strong coupling constant and f^{abc} structure constants of SU(3). In this case the covariant derivative is expressed as $D_\mu = \partial_\mu + ig_s T_a A^a_\mu$.

The full QCD Lagrangian is then given by

$$\mathscr{L}_{\text{QCD}} = -\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a + q_k(i\gamma^\mu D_\mu - m_k)q_k, \tag{1.8}$$

where the cubic and quartic terms of the gauge fields A^μ_a describe the gluon's self interaction.

In Quantum Field Theory, interactions between particles are influenced not only by the particles themselves but also by the continuous creation and annihilation of short-lived virtual

particles in the quantum vacuum. These off-shell fluctuations effectively "screen" or "anti-screen" the fundamental charges involved in the interaction, depending on the nature of the force. As a result, the coupling constants that characterize the interaction strengths are not fixed quantities but instead vary with the energy scale at which the process is probed.

This dependence on the energy scale is described by the process of renormalization, which systematically accounts for the effects of quantum fluctuations. The outcome is that each interaction in the Standard Model is governed by a running coupling constant, whose value changes according to the renormalization group equations.

In QCD, the strong coupling constant g_s also exhibits this running behavior. At low energy scales, corresponding to large distances, the coupling becomes strong, giving rise to the phenomenon of color confinement. In this regime, quarks and gluons cannot be observed as free particles, since any attempt to separate color charges requires so much energy that new quark–antiquark pairs are spontaneously created.

Conversely, at high energy scales, or equivalently at short distances, the strong interaction becomes weaker. This property, known as asymptotic freedom, allows quarks to behave almost as free particles inside hadrons and permits the use of perturbative techniques in theoretical calculations.

Because of confinement, quarks and gluons cannot exist as free particles; instead, they form color-neutral bound states known as *hadrons*. These composite particles – such as protons, neutrons, and mesons – are invariant under local SU(3) transformations, since their total color charge cancels, ensuring that all observable particles are color singlets.

The running of the strong coupling constant g_s is particularly relevant in the context of the research presented in this thesis. It underlies the formation of hadronic jets, which are the observable manifestations of quarks and gluons produced in high-energy collisions. The energy dependence of g_s allows high-energy proton–proton interactions at colliders such as the LHC to be interpreted in terms of the scattering of the protons' constituent partons. This framework provides the theoretical foundation for understanding the complex hadronic final states observed in experimental analyses.

1.1.3 The Electroweak Sector

The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam (GWS) model of electroweak interaction. The Lagragian is specified by requiring invariance under local $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge transformations, with $\mathrm{SU}(2)_L$ known as weak *isospin* and $\mathrm{U}(1)_Y$ as weak *hypercharge*.

The $\mathrm{SU}(2)_L$ generators are $T_a = \sigma_a/2$, with σ_a the three Pauli spin matrices, while the generator of $\mathrm{U}(1)_Y$ is the scalar $Y = 1/2$. Each of these generators correspond to a gauge bosonic field: $W_{1,2,3}$ and B, which, after the symmetry breaking (discussed in Section 1.2.1), give rise to the massive $W^{\pm}$ and Z vector bosons and the massless photon, defined as:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W_1 \mp iW_2), \tag{1.9}$$

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} cos\theta_W & sin\theta_W \\ -sin\theta_W & cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \tag{1.10}$$

where θ_W is the Weinberg angle which relates the electromagnetic and weak coupling constants through the following relation:

$$e = gsin\theta_W = g'cos\theta_W. \tag{1.11}$$

The massive vector bosons are the carriers of the weak force, and due to their mass the weak force has a short range and so it appear weak even though its intrinsic strength is comparable to that of QED.

The electroweak charge operator Q can be expressed as a combination of the third $\mathrm{SU}(3)_L$ generator and the $\mathrm{U}(1)_Y$ one as in

$$Q = T_3 + Y \ . \tag{1.12}$$

In the 1950s, the violation of parity conservation was discovered, showing that the weak interaction does not behave symmetrically under mirror reflection. This fundamental property makes the weak force unique among the interactions of the Standard Model. Only left-handed fermions and right-handed anti-fermions participate in weak processes, while their opposite chiral counterparts do not couple to the weak gauge bosons. As a consequence, and since neutrinos interact solely via the weak force, the Standard Model includes only left-handed neutrinos and right-handed anti-neutrinos, with no right-handed neutrino states.

1.2 The Brout-Englert-Higgs Mechanism

The Brout-Englert-Higgs mechanism (hereafter just "Higgs mechanism") provides the theoretical framework through which elementary particles acquire mass within the Standard Model, while preserving the gauge invariance of the underlying theory.

Before symmetry breaking, all gauge bosons and fermions are massless. In the unbroken electroweak theory, explicit mass terms cannot be introduced, as they would couple left- and right-handed components of the fermion fields, thereby violating the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge symmetry. Similarly, mass terms for the gauge bosons would explicitly break local gauge invariance by introducing longitudinal components incompatible with the gauge structure of the theory.

To resolve this, an additional complex scalar field, denoted by ϕ and referred to as the Higgs field, is introduced. Its potential is constructed such that the field acquires a non-zero vacuum expectation value (VEV), spontaneously breaking the electroweak symmetry. This spontaneous symmetry breaking endows the gauge bosons with mass, as discussed in Section 1.2.1, while the same mechanism also generates fermion masses through Yukawa interactions, as discussed in Section 1.2.2.

1.2.1 Electroweak Symmetry Breaking

The concept of spontaneous symmetry breaking (SSB) lies at the heart of the Higgs mechanism and the electroweak theory. It describes a situation in which the underlying Lagrangian of a system is invariant under a given symmetry group, but the ground state (vacuum) of the system is not. As a result, the symmetry of the equations of motion is not reflected in the vacuum configuration, leading to nontrivial physical consequences such as the emergence of massive gauge bosons and the appearance of scalar excitations.

To implement spontaneous symmetry breaking (SSB) within the Standard Model, a complex scalar doublet field ϕ is introduced. This field transforms as a doublet under $\mathrm{SU}(2)_L$ and carries hypercharge $Y = +1$ under $\mathrm{U}(1)_Y$. The presence of this field also allows for the inclusion of a symmetry-preserving term in the Standard Model Lagrangian, denoted by $\mathscr{L}_{\text{Yukawa}}$, which introduces mass terms for all fermionic fields in the Standard Model, except for the neutrinos.

The Higgs-sector part of the Standard Model Lagrangian is

$$\mathscr{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D_\mu \phi) - V(\phi) - \mathscr{L}_{\text{Yukawa}}, \tag{1.13}$$

where the covariant derivative D_μ acting on the scalar doublet, the one associated with $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ local gauge transformations, is

$$D_\mu \phi = \left(\partial_\mu - ig \frac{\tau^a}{2} W_\mu^a - ig' \frac{Y}{2} B_\mu \right) \phi, \tag{1.14}$$

.

The new introduced field ϕ is expressed as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.15}$$

with $\phi_{1,2,3,4}$ being real scalar field.

The potential term $V(\phi)$, invariant under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ transformations of the field, is defined as

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{1.16}$$

with $-\mu^2$ and λ constants. As can be observed, the $V(\phi)$ term contains both a quadratic and a quartic term in the scalar field ϕ.

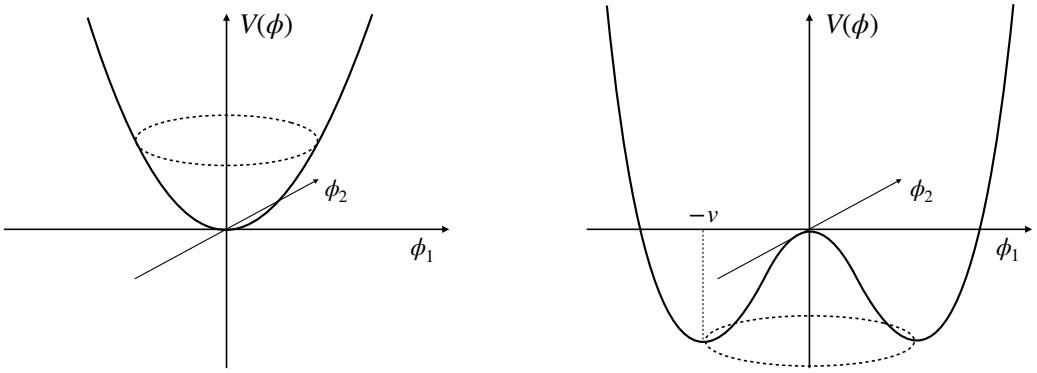


Figure 1.2: Higgs potential shape with $\mu^2 < 0$ (left) and with $\mu^2 > 0$ (right). The second scenario presents a continuum of equivalent minima where $|\phi| = v^2$.

The quadratic term, proportional to $-\mu^2$, determines whether the potential exhibits spontaneous symmetry breaking: as depicted in Figure 1.2, if $\mu^2 < 0$ the potential exhibit a unique minimum in $\phi = 0$, so that each excitation of the ϕ field in any direction in the (ϕ_1, ϕ_2) plane yields identical results, and in this case the quartic term is interpreted as a mass term for the scalar field. On the other hand, for $\mu^2 > 0$, the potential acquires the characteristic *Mexican-hat* shape which exhibit a degenerate continuum of minima around which the field is modeled as perturbations, leading to a situation in which excitations of the field in radial or tangential directions yields to different results, i.e. the symmetry is *broken*.

The vacuum expectation value (VEV) is the expected value of the field ϕ which minimizes the potential $V(\phi)$. A minimization of the potential lead sto a VEV ϕ_0 such that $v = \langle 0|\phi_0|0\rangle$ with value $\phi_0^2 = -\mu^2/\lambda = v^2$.

The value of the VEV is measured experimentally considering that the new introduced kinetic value for the ϕ field produces mass terms for the weak gauge bosons as

$$m_W = \frac{1}{2} g v, \qquad m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}, \tag{1.17}$$

with g and g' coupling constants of the gauge groups $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ respectively.

Taking this into account, and considering that, at tree level in the SM there is the valid relation

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2} \tag{1.18}$$

in which the Fermi constant G_F is measured precisely from muon decay and m_W can be expressed as a function of the VEV. In this way we obtain that $v = (\sqrt{2}G_F)^{-1/2}$, which leads to $v \simeq 246$ GeV.

The shape potential in Figure 1.2 shows a degeneracy in the direction pointed by the complex doublet ϕ, and as all the different vacuum states minimize the potential yielding identical physics, one can arbitrarily choose the state to lie along the second component of the doublet.

Applying the relation in Equation 1.12 shows that this choice is invariant under the charge operator, indicating the presence of an unbroken symmetry subgroup $\mathrm{U}(1)_Q$ under which the ground state is invariant.

Before symmetry breaking, the Higgs doublet can be parametrized in terms of four real scalar fields as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\,\phi_2(x) \\ v + h(x) + i\,\phi_3(x) \end{pmatrix}, \tag{1.19}$$

where $h(x)$ represents the physical excitation of the field around the vacuum expectation value v, while $\phi_{1,2,3}(x)$ correspond to three massless scalar fields, known as *would-be Goldstone bosons*. These fields arise due to the spontaneous breaking of the continuous $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ symmetry.

Through an appropriate local gauge transformation, one can remove these unphysical degrees of freedom. This particular choice of gauge is called the *unitary gauge*, in which the Higgs doublet takes the simpler form

$$\phi_{\text{unitary}}(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.20}$$

In this gauge, the Goldstone bosons $\phi_{1,2,3}$ are absorbed into the longitudinal components of the massive vector bosons $W^\pm$ and Z, providing the necessary third polarization state for each. The remaining real scalar field $h(x)$ corresponds to a physical, observable particle: the Higgs boson.

Rewriting the Higgs kinetic term inside the Lagrangian including the definition of the ϕ field, and identifying physical fields from the quadratic terms of the linear combination of the unphysical ones, the physical $W_\mu^\pm$, Z_μ and A_μ fields can be written in terms of the original A_μ^a and B_μ fields. This gives

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2), \qquad \begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix} = \begin{bmatrix} cos\theta_W & sin\theta_W \\ -sin\theta_W & cos\theta_W \end{bmatrix} \begin{bmatrix} B_\mu \\ A_\mu^3 \end{bmatrix} \tag{1.21}$$

with θ_W weak mixing angle defined through the coupling constants as $\theta_W = g/\sqrt{g^2 + g'^2}$.

The corresponding mass terms of the electroweak gauge bosons and the interaction terms between these bosons and the Higgs field can be read off as:

$$m_W = \frac{1}{2}gv, \qquad m_Z = \frac{m_W}{cos\theta_W}, \qquad m_\gamma = 0. \tag{1.22}$$

Thus, in the unitary gauge, the spontaneous symmetry breaking mechanism manifests as the generation of massive $W^\pm$ and Z bosons, while the photon remains massless, corresponding to the unbroken $\mathrm{U}(1)_{\text{em}}$ subgroup.

In addition to generating masses for the gauge bosons, the spontaneous symmetry breaking also affects the scalar potential itself. By substituting the unitary gauge expression of the Higgs doublet $\phi_{\text{unitary}}(x)$ into the Higgs potential

$$\begin{aligned} V(h) &= -\frac{\mu^2}{2}(v+h)^2 + \frac{\lambda}{4}(v+h)^4 \\ &= \text{const.} + \frac{1}{2}(2\lambda v^2)\, h^2 + \lambda v\, h^3 + \frac{\lambda}{4}\, h^4. \end{aligned} \tag{1.23}$$

The quadratic term in this expansion defines the mass of the physical Higgs boson,

$$m_h^2 = 2\lambda v^2, \tag{1.24}$$

while the cubic and quartic terms describe, respectively, the trilinear and quartic self-interactions of the Higgs field. The parameter λ thus governs both the strength of these self-couplings and the value of the Higgs mass. Using the measured values $m_h \simeq 125$ GeV and $v \simeq 246$ GeV, one finds

$$\lambda = \frac{m_h^2}{2v^2} \simeq 0.13, \tag{1.25}$$

indicating that the Higgs field is only weakly self-interacting within the Standard Model.

In summary, spontaneous symmetry breaking occurs when the Higgs field acquires a nonzero vacuum expectation value, causing the underlying $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge symmetry to be hidden in the low-energy spectrum. As a result, the weak gauge bosons and fermions acquire mass, while the photon remains massless, and the Higgs field manifests as a single physical scalar particle whose mass and self-interactions are directly determined by the shape of the scalar potential. This mechanism elegantly links the abstract symmetry of the theory to the concrete properties of the observed particles in nature.

1.2.2 Fermionic Yukawa Coupling

The addition of the fermions masses into the theory breaks the gauge invariance under $\mathrm{SU}(2)_L$ gauge transformations. Symmetry conserving mass terms for the fermions arise from the Yukawa coupling term between the fermions and the Higgs field when the system undergoes SSB.

In this way their masses are determined by their couplings with the Higgs field and by the Higgs field VEV, which sets the basic mass scale of the theory.

Recalling that the Higgs field transform as an SU(2) doublet with $Y = 1/2$, as the left-handed fermion field ϕ_L, and that the right-handed fermion field ϕ_R transforms as an SU(2) singles, the re-normalizable and gauge invariant coupling between a fermionic field ψ and the scalar Higgs field ϕ can be written as

$$\mathscr{L}_{\text{Yukawa}} = -y_f(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L), \tag{1.26}$$

with y_f Higgs-fermion Yukawa coupling. If considering only the first generation leptons, $\psi_L = (\nu_L, e_L)$ and $\psi_R = e_R$, after the spontaneous symmetry breaking, i.e. when expressing the field ϕ in terms of the VEV and the Higgs field $h(x)$, the Yukawa Lagrangian term become

$$\mathscr{L}_{\text{Yukawa}} = -\frac{v y_e}{\sqrt{2}}\bar{e}e - \frac{y_e}{\sqrt{2}}\bar{e}he. \tag{1.27}$$

Here, the VEV component of the ϕ field appears into the quadratic electron field term, that can therefore be identified as the electron mass term $m_e = v y_e/\sqrt{2}$. No masses can be addressed to the neutrino field due to the absence of right-handed neutrinos. The higgs field h,

appearing in the second term, make it representing the interaction between the electron field e and the Higgs one.

In the same way mass is generated also for the other charged lepton generations and down-type quarks, with $\psi_L = (u_L, d_L)$ and $\psi_R = d_R$ for the first quark generation. For the up-type quarks also the mass in generated with the same process, but in this case their Yukawa coupling is with the conjugate field to ϕ, $\tilde{\phi}$, constructed as

$$\tilde{\phi} = i\sigma_2\phi^\dagger = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1^\dagger \\ \phi_2^\dagger \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} v+h \\ 0 \end{bmatrix} \tag{1.28}$$

and transforms in the same way as ϕ.

In the Standard Model, the Yukawa interactions generally couple quarks from different generations, leading to non-diagonal mass matrices after spontaneous symmetry breaking. The quark fields that diagonalize these mass matrices define the *mass eigenstates* q, which correspond to the physical particles observed in experiments. However, the electroweak interactions act on the quark fields in the *weak eigenstate* basis $\tilde{q}$, which are linear combinations of the mass eigenstates.

This misalignment between the mass and weak bases implies that charged-current weak interactions can change quark flavour: a quark of a given mass can interact with a W boson and transform into a quark of a different generation. The strength of these flavour-changing interactions is encoded in the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM}, defined as

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}, \tag{1.29}$$

that is a unitary matrix whose elements V_{ij} quantify the coupling between the up-type quark of generation i and the down-type quark of generation j. The CKM matrix not only governs the pattern of quark mixing and flavour-changing weak decays, but also introduces a complex phase that is responsible for CP violation in the quark sector.

1.3 Higgs Sector Phenomenology at LHC

The properties of the Higgs boson are directly determined by its role in the Standard Model. In particular, the Higgs couples to other particles proportionally to their masses: the Yukawa couplings to fermions are proportional to the fermion mass, while its couplings to gauge bosons are proportional to the square of the boson masses. This relationship dictates both the production mechanisms of the Higgs at the LHC and its decay patterns. Understanding these connections is crucial for interpreting experimental measurements and testing the consistency of the Standard Model.

The production cross sections and decay branching ratios of the Higgs boson are therefore direct probes of its couplings. The strength of each production mode depends on the corresponding Higgs coupling: for example, gluon-gluon fusion occurs via a top-quark loop and is sensitive to the top Yukawa coupling, whereas vector boson fusion and associated production with a vector boson probe the Higgs–gauge boson couplings. Similarly, the decay rates into fermion pairs are proportional to the square of the corresponding Yukawa coupling, while decays into gauge bosons depend on the Higgs–vector couplings.

Figure 1.3 summarizes the Standard Model predictions for Higgs production cross sections and decay branching ratios at $\sqrt{s} = 13$ TeV as a function of the Higgs mass.

At leading order in QCD, Higgs boson production occurs mainly through four modes, showed in Figure 1.4: the gluon-gluon fusion (ggF), which is predoninantly mediated by a virtual top

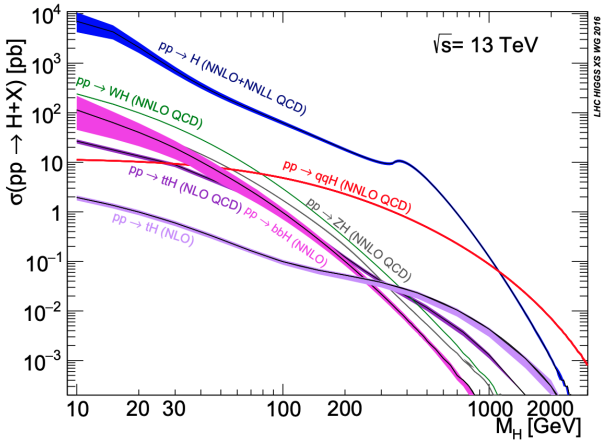


Figure 1.3: Standard Model predictions for Higgs production cross sections as a function of the Higgs mass at $\sqrt{s} = 13$ TeV. At $M_H = 125$ GeV the main production modes of the Higgs boson are via gluon-gluon fusion, vector boson fusion, associated production with vector bosons and top quark fusion [3].

quark loop, the vector-boson fusion (VBF), in which a pair of W or Z bosons fuse to produce a Higgs after being radiated by two quarks, the VH production (or Higgs-strahlung), in which the Higgs boson is produced in association to a vector boson, and top quark fusion ($t\bar{t}H$ or more generarly $q\bar{q}H$), in which two gluons radiates each a quark-antiquark pair and the quark from each pair fuses to produce a Higgs boson.

Despite its high production rate, ggF presents significant experimental challenges: the Higgs is typically produced without additional high-momentum objects in the final state, and its decays, particularly into $H \to b\bar{b}$, are buried under overwhelming QCD multi-jet backgrounds. Consequently, although ggF dominates in terms of raw production, it is not the most favorable channel for observing certain decay modes such as $H \to b\bar{b}$.

In contrast, associated production with a vector boson provides a cleaner signature: the leptonic decay products of the W or Z boson serve as triggers and help to reduce backgrounds, making VH the preferred channel for studying the Higgs-bottom Yukawa coupling.

Once produced, the Higgs boson decays almost instantaneously, with a lifetime on the order of $10^{-22}s$. The pattern of its decay channels is entirely determined by the structure of its couplings within the Standard Model. Since the Higgs field interacts with fermions through Yukawa terms proportional to their masses, the partial decay width into a fermion pair scales as $\Gamma(H \to f\bar{f}) \propto m_f^2 m_H$, where m_f and m_H are respectively the fermion and Higgs masses. As a result, heavier fermions couple more strongly to the Higgs and dominate its decay pattern whenever kinematically allowed. Similarly, the decays into massive vector bosons, $H \to WW$ and $H \to ZZ$, depend on the Higgs–boson couplings, which are proportional to the squared masses of the gauge bosons, m_V^2.

For a Higgs boson with a mass of $m_H = 125$ GeV, the dominant decay mode is into a botto-quark pair, accounting for roughly 58% of the total width. Decays into WW^* and ZZ^* are also significant, where one of the vector bosons is off-shell due to kinematic constraints. Other relevant channels include $H \to \tau^+\tau^-$ and $H \to c\bar{c}$, reflecting the hierarchy of the fermion masses. Loop-induced processes, such as $H \to \gamma\gamma$ and $H \to Z\gamma$, occur through virtual top-quark or $W-$boson loops and have much smaller branching ratios, but play a crucial role experimentally thanks to their clean final states.

The complete picture of Higgs decays, shown in Figure 1.5, reveals the smooth dependence of the branching fractions on the Higgs mass: at low masses, decays into fermions dominate, while at higher masses ($m_H > 160$ GeV), the openning of the WW and ZZ channels rapidly

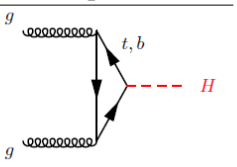
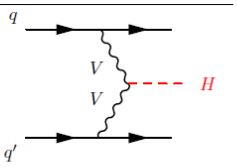
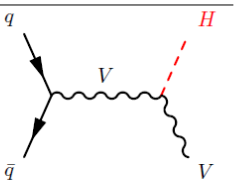
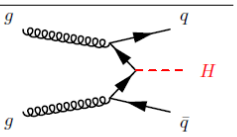
Production mode	LO diagram
ggF production	
VBF production	
VH production	
$q\bar{q}H$ production	

Figure 1.4: Diagrams for the four main Higgs production mechanisms at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV.

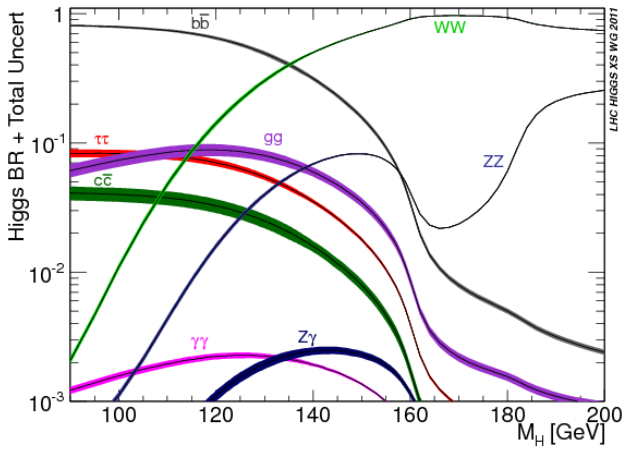


Figure 1.5: Standard Model predictions for the Higgs boson branching ratios as a function of the Higgs mass. The values at $m_H = 125$ GeV correspond to the experimentally observed Higgs boson [4].

suppresses fermionic modes.

In summary, the study of Higgs boson production and decay provides a powerful means to probe the structure of the Standard Model and to test the mechanism of electroweak symmetry breaking through precise measurements of the Higgs couplings. Among the various production and decay channels, the associated production with a vector boson followed by the decay $H \to b\bar{b}$(commonly referred to as the VHbb channel) plays a particularly important role. Although its production rate is smaller than that of gluon-gluon fusion, the presence of a leptonic W or Z decay in the final state offers a clean experimental signature and significantly reduces the overwhelming QCD background that affects purely hadronic Higgs decays.

In this thesis, special attention is devoted to this process: the VHbb channel in the boosted regime is studied in detail, providing direct sensitivity to the Higgs coupling to bottom quarks and serving as a crucial test of the Yukawa sector of the Standard Model. The methodology, event reconstruction techniques, and results of this measurement are discussed extensively in Chapter 5, where the analysis strategy and its implications for Higgs physics at the LHC are presented.

Chapter 2

The ATLAS experiment at the LHC

2.1 The Large Hadron Collider

In December 1994, the CERN (Conseil Européen pour la Recherche Nucléaire, or European Organization for Nuclear Research) Council in Geneva approved the construction of the Large Hadron Collider (LHC), describing it as the "primary research tool at the energy frontier of particle physics [5]. Once approved, it has become the largest and most powerful particle accelerator ever built by humans.

The LHC was constructed between 1998 and 2008 within the tunnel that previously housed the Large Electron-Positron (LEP) collider. That earlier facility played a key role in discovering and testing many foundational aspects of modern physics with impressive precision. The LHC itself is a 27-kilometer ring that accelerates protons and heavy ions to TeV-scale energies, which then collide at four different points along the ring.

Each of these interaction points hosts a detector designed to measure and analyze the products of the proton-proton collisions and their subsequent decays. The experiments are equipped with diverse architectures and technologies, allowing the LHC to explore a wide range of phenomena across different fields of high-energy physics.

The main detectors are ATLAS, CMS, LHCb, and ALICE. ATLAS and CMS are versatile detectors capable of observing signals both within the Standard Model and beyond, and for this reason they are usually called *multi-purpose* detectors. LHCb specializes in heavy flavor physics and rare decays, while ALICE focuses on studying heavy-ion collisions to understand the properties of strongly interacting matter under extreme conditions, including the quark-gluon plasma.

During my PhD I worked in the ATLAS Collaboration, in which the results of proton-proton collisions are collected through several steps which will be explained in the upcoming section.

Prior to data taking, protons are accelerated through a sequence of stages that gradually raise their energy to the value required for injection into the LHC ring as outlined in Figure 2.1.

Everything starts from hydrogen gas (H_2), in which two hydrogen atoms are bounded in a single molecule, then, using a ion source the gas is ionized in order to strip away electrons from protons in the hydrogen atoms obtaining free protons (H^+ ions). Then, an electromagnetic field is used for extracting the protons and pulling them into a well defined, narrow beam, subsequently focused and shaped using magnets. Then the beam is injected in the first accelerator (initially LINAC2 and replaced by LINAC4 in 2020), in which, passing through radio-frequency (RF) cavities, the protons gain kinetic energy while being focused by quadrupoles, passing from a var low energy (few keV) up to about 50 MeV. The accelerated protons are then extracted using magnetic or electrostatic deflectors to direct the beam into the transfer lines leading to the next accelerator, the Proton Synchrotron Boosted (PSB).

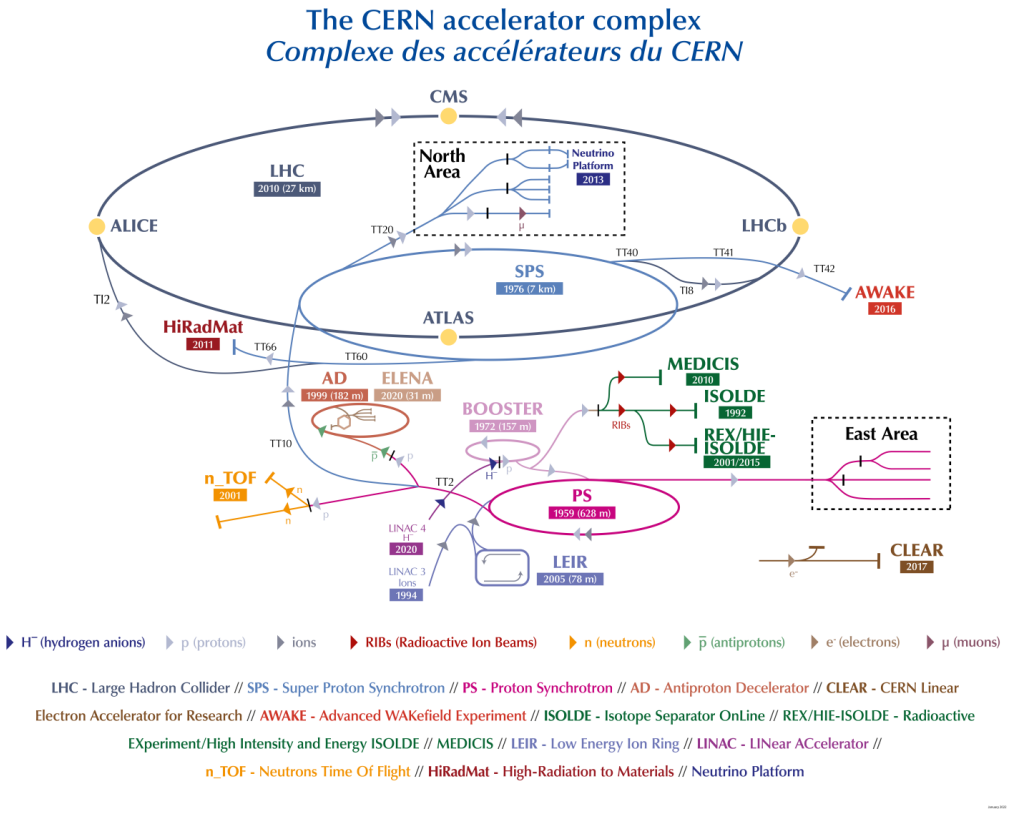


Figure 2.1: The CERN accelerator complex [6]

Inside the PSB, quadrupole magnets continue to focus the proton beam, maintaining a narrow and stable bundle of particles while accelerating it up to 1.4 GeV. At this stage the proton beam is divided into bunches (grouping the protons into tight packets synchronized with the RF system), which are subsequently injected in the Proton Synchrotron (PS).

The PS serves as a critical intermediate step, preparing the beam for transfer to larger accelerators and bringing the energy of the protons from 1.4 GeV up to approximately 25 GeV.

Then, the Super Proton Synchrotron (SPS) accelerates the bunch of protons to an energy of 450 GeV and finally the beam arrive to the LHC ring. In it, thanks to a doublet injection system, two beam pipes are used for making protons circulate in opposite directions into the ring.

Once injected, the protons are rapidly accelerated over many laps of the 27 km ring. The magnetic fields of the LHC's dipole magnets reach values of 8.3T, ramping up gradually the beams' energy up to 6.8 TeV, while higher order magnets are using for additional shaping and focusing of the beams.

At this point, in stable beam conditions, the nominal time-spacing between two bunches is of 25 ns.

Before describing the ATLAS experiment, it is important to introduce the key quantities that characterize the operation of a particle collider: the instantaneous luminosity $\mathcal{L}_{\text{inst}}$ and the center-of-mass energy $\sqrt{s}$.

The **instantaneous luminosity** is a measure of how frequently particles collide within a collider at any given moment. It essentially quantifies the potential number of collision events happening per unit area per second. The higher the luminosity, the more collisions occur each second, which is particularly important for studying rare processes in particle physics. It is defined as the number particles crossing per unit of area and time, and relates the cross-section

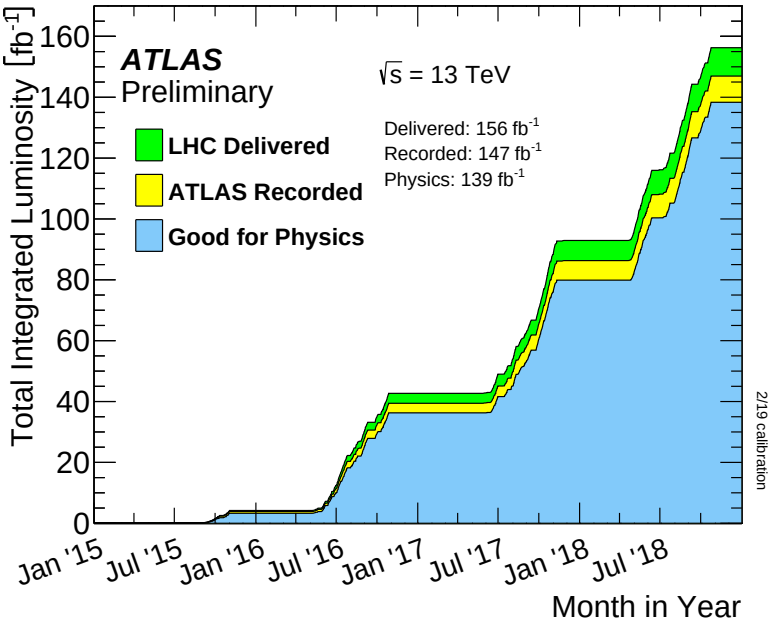


Figure 2.2: ATLAS Total Integrated Luminosity and Data Quality achieved over full Run2 [7]

(σ) of a given process to the rate of events of such process dN/dt by:

$$\frac{dN}{dt} := \mathcal{L}_{inst} \times \sigma, \tag{2.1}$$

where the cross-section of a process (expressed in cm^2) is proportional to the probability of the process to occur and depends on both the physics process considered and the energy of the collision (as some resonances cannot appear or can be suppressed at certain energy values).

Cross-section is usually measured in barns (b), defined such that $1b = 10^{-24}cm^2$, and consequently the instantaneous luminosity $\mathcal{L}_{inst}$ can be expressed both in $cm^{-2}s^{-1}$ and $fb^{-1}s^{-1}$.

Is important to notice that the instantaneous luminosity is an intrinsic property of the accelerator and depends on its settings, in fact, it can also be expressed in terms of the collider's parameters as:

$$\mathcal{L}_{inst} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y}. \tag{2.2}$$

Here, N_1 and N_2 are the number of protons in the two colliding bunches, f is the revolution frequency of the collider, N_b is the number of bunches per beam and σ_x and σ_y represent the widths of the beam density distributions along the x and y directions.

Integrating $\mathcal{L}_\rangle\backslash\int\sqcup$ over a time period T, the integrated luminosity $\mathcal{L}_{int} = \int_0^T \mathcal{L}_{inst}(T)dt$ is obtained. This quantity provides a measure of the total data collected by the experiment, and is crucial because it correlates directly with the statistical power of the experimental results. Also, having the cross-section for a given process, the integrated luminosity allows to calculate the total number N of such process events produced at the collider by:

$$N = \mathcal{L}_{int} \times \sigma. \tag{2.3}$$

On the other hand, the **center of mass energy** $\sqrt{s}$ refers to the energy that the colliding particles have in their center-of-mass reference frame. It essentially represents the maximum energy that can be used to produce new particles during the collision, and so defines the energy frontier that the collider can explore.

The operational periods of LHC can be divided, until now, in three data taking phases. In the **Run 1** (2009-2012), the LHC operated at $\sqrt{s} = 7$ and 8 TeV and where the ATLAS and CMS experiments collected 25 $fb^{-}1$ good quality data (taken under good operating conditions). After it, a first long shut down period (LS1) followed for the upgrade and for repairing such detector altered by the radiation. Then, in the **Run 2** (2015-2018) at $\sqrt{s} = 13$ TeV the two experiments collected 139 $fb^{-}1$ each (Figure 2.2). The data recorded during this period form the basis of the analysis presented in this thesis. Finally, after a second long shut down (LS2), in 2022 the **Run 3** started, reaching in July the highest energy ever achieved at a collider of $\sqrt{s} = 13.4$ TeV, , and is expected to ultimately triple the size of the collected dataset.

In high-energy collider experiments, the primary events studied are head-on proton-proton collisions in a single bunch crossing, which mainly result in inelastic interactions between the protons' constituent partons (quarks and gluons). However, to accurately interpret the collected data, experiments have to deal with the additional soft interactions that accompany these hard scatterings and occur simultaneously. These supplementary interactions are collectively known as 'pile-up'. Pile-up interactions can significantly affect the measurement and analysis of collision events and are typically categorized into two types: in-time pile-up, which involves soft interactions that take place within the same bunch crossing as the primary hard collision; and out-of-time pile-up, which comprises soft interactions originating from neighboring bunch crossings that occur in subsequent or preceding beam crossings but can still influence detector signals due to detector response times and data acquisition effects.

Pile-up for each data taking period [7] is quantified by the average number of interactions per bunch crossing $\langle\mu\rangle$, which is computed using the average instantaneous luminosity per bunch crossing $\mathcal{L}_{bunch}$ and the inelastic scattering cross-section for the proton-proton collisions σ_{inel}, together with the number of bunch N_b and the revolution frequency f as

$$\langle\mu\rangle = \frac{\mathcal{L}_{bunch}\sigma_{inel}}{N_b f}. \tag{2.4}$$

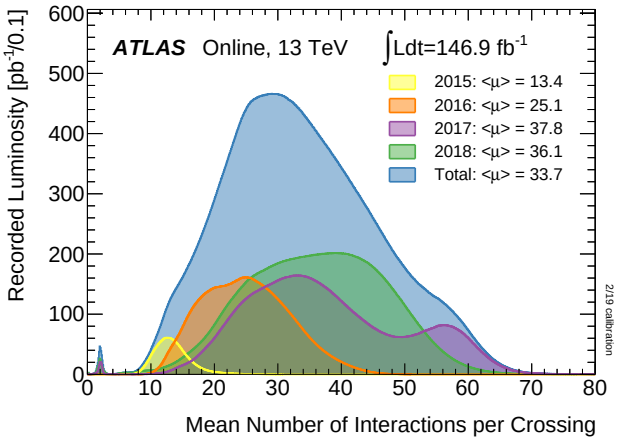


Figure 2.3: Average number of pile-up interactions per event observed in the ATLAS experiment during the Run-2 of LHC. [7]

The $< \mu >$ parameter is therefore fundamental for correcting for the pile-up effects in data analysis and is used as per-year correction on the simulated samples to agree with the data. The distribution of its observed values in the ATLAS detector in the different years of the Run-2 is showcased in Figure 2.3, when it varied between 13 and 38 over the years for a final mean value of 33.3.

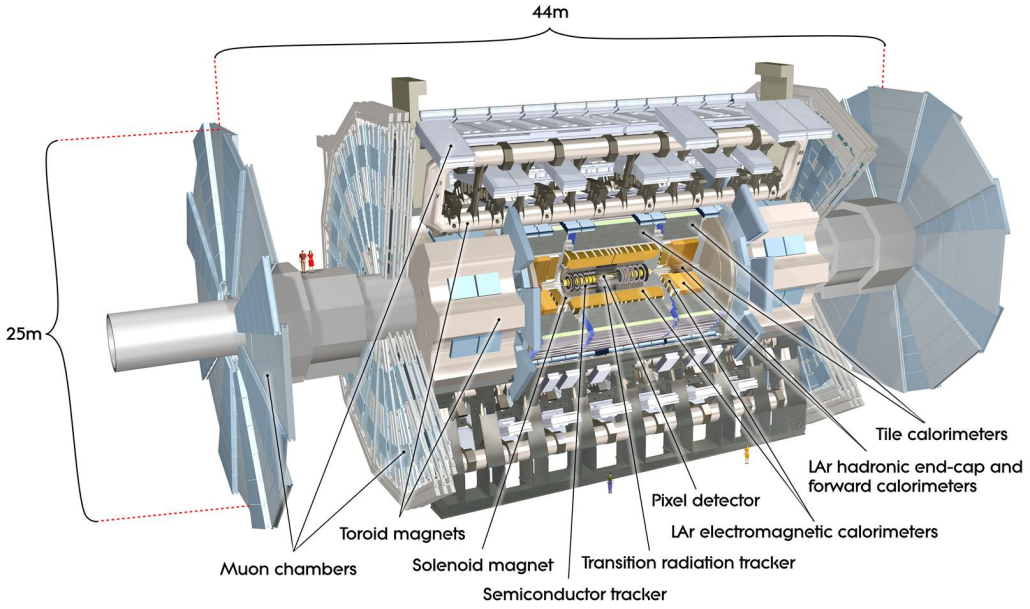


Figure 2.4: ATLAS detector scheme with its subsystems

2.2 The ATLAS Experiment during Run2

The ATLAS (A Toroidal LHC ApparatuS) experiment, showed in Figure 2.4, is a multi-purpose particle detector designed to explore a wide range of physics phenomena. With its 25 meters of diameter, its 44 meters of length and its 7000 tons of weight, it is the larger detector within the LHC complex. It has a cylindrical geometry and comprises multiple layers of sub-detectors, each optimized to measure different properties of the particles produced. This layered structure enables the reconstruction of particle trajectories and energies across nearly the entire solid angle, approximately (4π).

To ensure comprehensive particle detection coverage, the detector is divided into two main components: the *barrel*, which constitutes the central region of the cylindrical structure, and the *endcaps*, which cover the detector's respective ends, thereby providing full angular coverage.

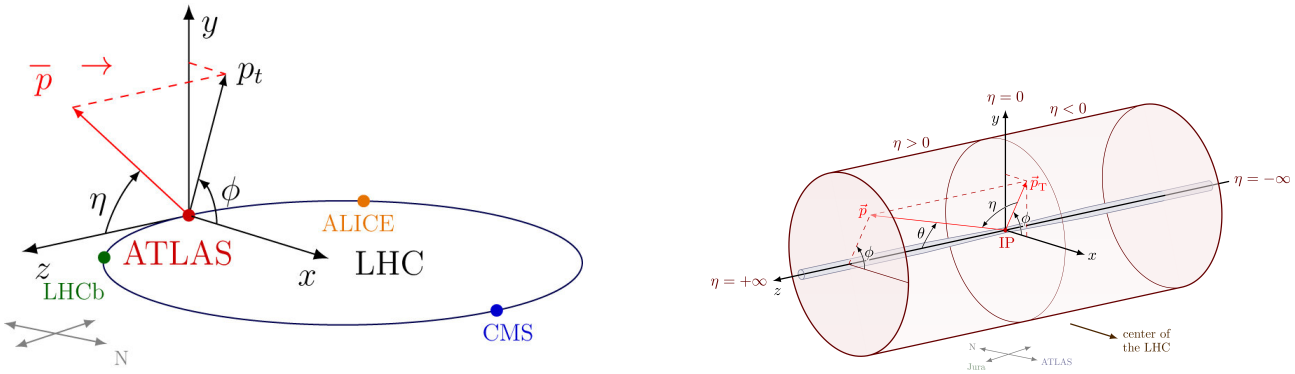
The ATLAS detector is composed of several specialized sub-detectors, each tailored to measure specific types of particle interactions and properties. These sub-detectors are arranged in a layered configuration around the collision point, working in concert to provide detailed information about the particles produced in high-energy collisions. The primary sub-detectors include the inner tracking system, the calorimeters, and the muon spectrometer, each playing a crucial role in reconstructing the trajectories, energies, and identities of the particles with high precision.

Before delving into the detailed descriptions of each sub-detector, it is essential to establish a clear understanding of the ATLAS coordinate system.

The **ATLAS Coordinate System**, with its origin placed in the interaction point (IP), refers to the beam-director as the z-axis; the x-axis is the one towards the center of the LHC ring and the y-axis points upwards from the IP. The xy-plane is commonly called transverse plane while the longitudinal one refers to any plane containing the z-axis.

In the transverse plane the polar coordinates are used, defining the θ angle as the one measured upwards from the beam axis ($\theta \in [0, \pi]$) and using the azimuthal ϕ one for spanning the xy-plane ($\phi \in [0, 2\pi[$).

In ATLAS the θ angle is commonly substituted with an alternative variable, the *rapidity*, which differences are Lorentz-invariant.



(a) (b)

Figure 2.5: ATLAS coordinate system with respect to LHC (a) and with respect to the detector itself (b) [8]

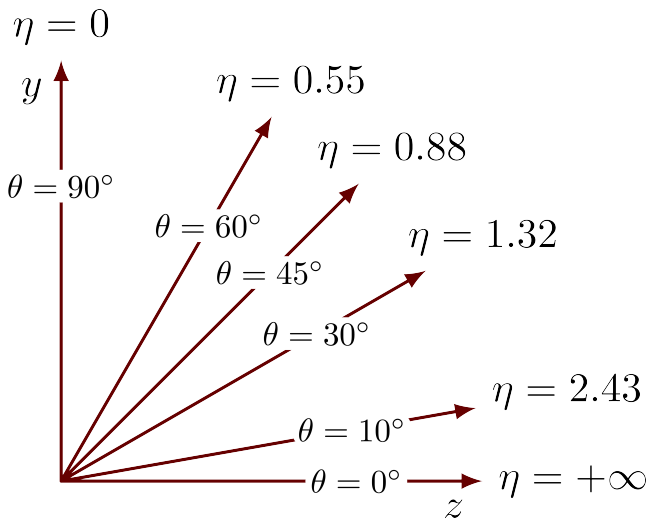


Figure 2.6: Pseudo-rapidity values as a function of the polar angle. $\eta = 0$ correspond to the transverse plane and $\eta = \pm\infty$ to the beam axis [8].

Rapidity is indicated with the y symbol and defined as

$$y = \frac{1}{2} ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2.5}$$

with p_z projection of the particle momentum along the z-axis and E the particle's energy, both extracted by the four-momentum of the particle $(E, \mathbf{p})$ with $\mathbf{p} = (p_x, p_y, p_z)$. For highly energetic particles (i.e. with $p >> m$), the $pseudo - rapidity$ η is considered as a good approximation of the rapidity y and gives the advantage of relying only on geometrical information instead of the particle properties like the momentum or the energy. It is defined as

$$\eta := -ln \left[tan \left(\frac{\theta}{2} \right) \right], \tag{2.6}$$

and represented as a function of the polar angle θ in Figure 2.6.

Another Lorentz invariant kinematic quantity commonly used is the ΔR, representing the angular distance between two objects in the detector and defined as

$$\Delta R = \sqrt{(\phi_1 - \phi_2)^2 + (\eta_1 - \eta_2)^2}. \tag{2.7}$$

Due to the unknown momenta of the partons along the $z-$axis, energies and momenta of the particles are commonly measured in the xy-plane (called *transverse* plane), as $p_T = |\mathbf{p}|sin(\theta)$ and $E_T = Esin(\theta)$.

A final comprehensive description of the coordinate system of the ATLAS detector is depicted in Figure 2.5.

Having established the ATLAS coordinate system, we now proceed to describe the various sub-detectors that comprise the detector and enable the measurement of particle properties throughout its different regions.

ATLAS Magnetic System

When a particle carrying electric charge enters an electric field, it experiences a force proportional to its charge and the field intensity, according to Coulomb's law. This results in an acceleration along the direction of the electric field. Conversely, when a charged particle traverses a magnetic field, the force is given by the Lorentz equation

$$\mathbf{F} = q\ (\mathbf{E} + \mathbf{v} \times \mathbf{B}), \tag{2.8}$$

with q particle's charge, v its velocity, and E and B respectively the electric and magnetic fields. In the absence of an electric field, the force is always orthogonal to the direction of motion, leading to a curved trajectory. For relativistic particles in a uniform magnetic field, this results in a helical path, with a circular projection in the plane perpendicular to the field and free propagation along the field direction.

The radius of curvature of the circular component can be related to the magnetic field strength and the particle's transverse momentum through the following expression:

$$p_T[GeV] = 0.3qB[T]r[m]. \tag{2.9}$$

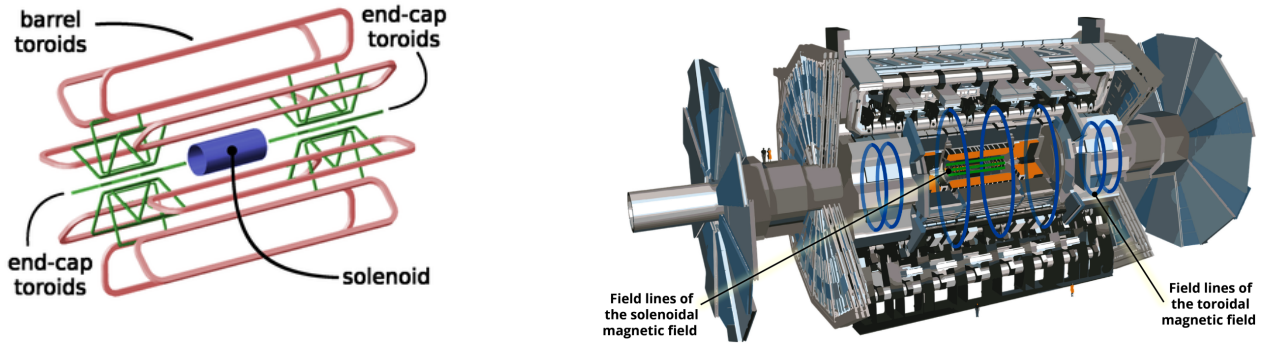
Thus, by measuring the deflection of a charged particle's trajectory, one can infer its momentum. The sign of the curvature also reveals the charge of the particle. In practical terms, tracking detectors embedded within magnetic fields provide simultaneous access to fundamental kinematic observables like the charge sign (looking at the bending direction), the momentum magnitude (from the radius) and the momentum components (by combining the curvature in the transverse plane with the straight-line component parallel to the field).

These principles underpin the design of modern collider experiments, where high precision in momentum reconstruction is required across a broad energy range.

The ATLAS magnet system, depicted in Figure 2.7 together with the direction of the generated field lines, employs a complex magnetic configuration to maximize its tracking and muon identification capabilities. It consists in two subsystems, the solenoidal and the toroidal magnet systems. The solenoidal system surrounds the inner tracking system providing a 2 T axial magnetic field parallel to the beam line, ensuring precise momentum measurements of charged particles emerging from the interaction point.

The toroidal system is placed beyond the calorimeters and is composed by a central barrel toroid and two end-cap toroids, each made of eight superconducting coils and generating respectively 0.5 T and 1 T magnetic fields predominantly orthogonal to the trajectories of muon escaping the calorimeters.

The combination of solenoidal and toroidal fields enables ATLAS to achieve excellent momentum resolution over a wide kinematic range. The solenoid ensures precise measurement of charged particle tracks in the inner detector, while the toroids provide robust muon momentum determination in the outer regions. Together, they form a magnetic system that is essential for the experiment's capability to identify particles, measure their properties, and ultimately explore phenomena at the energy frontier.



(a) (b)

Figure 2.7: ATLAS schematic magnets system with the solenoid and the barrel and end-cap toroid magnets (a) together with the magnetic field lines they produce (b)

.

2.2.1 Inner Detector and Tracking System

The ATLAS Inner Detector (ID), schematized in Figure 2.8, is the innermost sub-system of the ATLAS detector, the closer one to the LHC beam-pipe. It is responsible for identifying charged particles originating from the interaction point (IP) and, by analyzing the curvature of particle tracks within the solenoid magnetic field in which it is embedded, it enables precise measurements of their momentum.

The Inner Detector covers the region with $|\eta|$ <2.5 and consists in three sub-systems: the Pixel Detector, the SemiConductor Tracker (SCT) and the Transition Radiation Tracker (TRT), placed both in the barrel and in the end-cap regions. A cutaway view of the entire system is depicted in Figure 2.9.

Pixel Detector

The Pixel Detector is the innermost part of the ID, it is a silicon based detector composed of 4 high granular ϕ and z segmented cylindrical layers and 2 end-caps with 3 disc layers each. The barrel layers are placed between 33.25 and 122.5 mm around the beam pipe, with the full barrel+endcap detector system covering $|\eta|$ <2.5.

The Insertable B-Layer (IBL), introduced between Run-1 and Run-2 just before the innermost pixel layer, is composed of pixels with size of 50 μm in the $r-\phi$ direction and 250 μm in the z one and results in the highest granularity of the entire detector. Its primary purpose is to enhance the robustness of tracking capabilities and to accommodate the increased peak luminosity of Run-2. Moreover, the IBL plays a crucial role in the jet-flavour identification, an aspect explored in detail in Chapters 4 and 5, which utilize full Run-2 data.

With it, the impact parameter (IP) resolution, defined as the closest approach of a track to the collision point, is significantly improved, by about 40%, for low-p_T tracks ($p_T < 1$ GeV) [11].

The three remaining layers have a pixel size of 50 μm in the $r-\phi$ direction and 250 μm in the z one, resulting in an expected hit resolution of 8×100 μm in the $r-\phi$ direction and 40 μm and 115 μm in the $z-$direction respectively for the IBL and the three remaining layers. Charged particles leaves an average of 4 hits in the pixel detector.

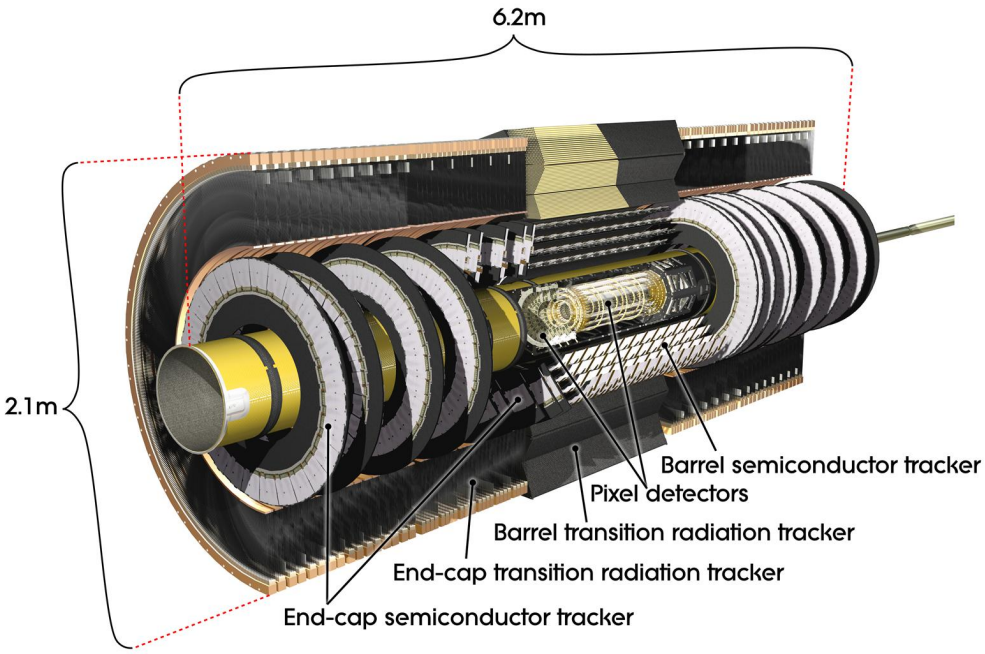


Figure 2.8: Three dimensional scheme of the ATLAS Inner Detector [9].

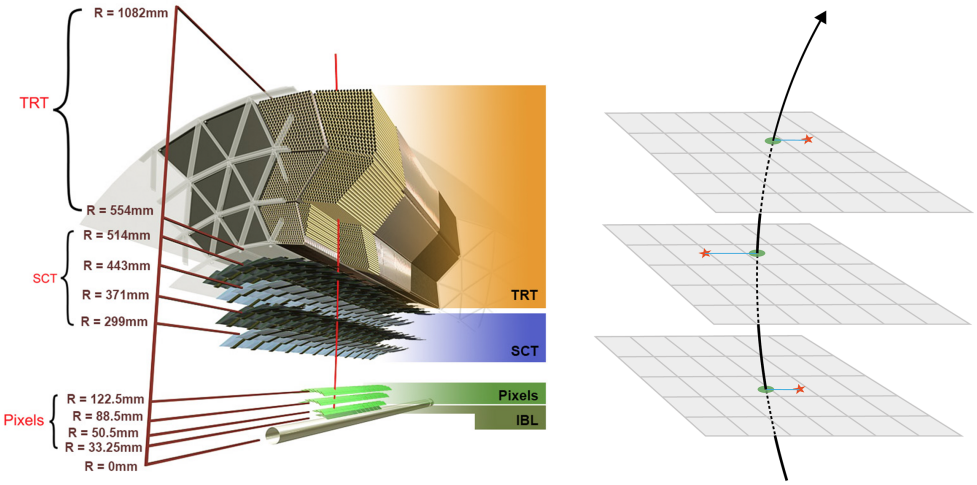


Figure 2.9: Transverse view of the ATLAS ID [10]

SemiConductor Tracker

The SemiConductor Tracker (SCT) is a precise silicon microstrip module based detector placed around the Pixel Detector. Its strips typically measure 80 μm in size, and each module contains two sensor layers tilted at a small stereo angle of 40 mrad, which allows for two-dimensional hit reconstruction. The barrel region, covering $|\eta|$ <1.4, consists of four cylindrical layers of these strip modules, while each end-cap, spanning 1.4 < $|\eta|$ <2.5, is made up of nine disc-shaped modules. The intrinsic spatial accuracy of the SCT is approximately 17 $\times 580 \mu m^2$ in the barrel and 580×17 μm^2 in the end-cap regions. Charged particles typically produce around eight hits within the SCT. When combined with measurements from the pixel detector, these hits provide valuable additional information that can be used to enhance jet flavor tagging, especially at high p_T.

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is the last layer of the inner detector. It is straw-tube tracked composed of drift tubes with a diameter of 4 mm filled with an Argon-based gas mixture. Each tube contains a thin tungsten wire of 31 μm placed in its center, functioning as a small proportional counter to collect charges generated by ionization caused by passing charged particles.

In the barrel region ($|\eta| < 1$) the TRT comprises 52544 straw-tubes of 1.5 meter length placed parallel to the beam axis. The end-cap region ($1 < |\eta| < 2$) contains 122880 straws each long 0.4 meters arranged perpendicularly to the beam axis.

The TRT also exploits the phenomenon of transition radiation: when highly relativistic particles traverse boundaries between different materials, such as polymer fibers in the barrel or foils in the end-cap, they emit X-ray photons if their Lorentz boost ($\gamma \sim E/m$) is sufficiently high.

By examining these signals, the TRT improves particle identification, especially in differentiating electrons from pions, since electrons, being lighter, attain higher γ at the same energy and are thus more likely to emit detectable transition radiation.

2.2.2 Electromagnetic and Hadronic calorimeters

Unlike trackers, calorimeters are designed to measure the energy of neutral and charged particles by absorbing them completely. They are positioned after the pixel detector and operate by forcing particles to deposit all their energy within their coverage (which provides near hermetic coverage in the azimuthal angle, ϕ, and a pseudorapidity range of $|\eta| < 4.9$).

Calorimeters are constructed from multiple layers of active and passive materials, as illustrated in Figure 2.10. The passive layers are typically composed of dense substances with high atomic numbers, serving to absorb and stop incoming particles, while the active layers detect energy deposits by recording ionization or scintillation light produced by particle-induced showers.

Two distinct types of calorimeters are used to optimize measurements for different particle types: the Electromagnetic Calorimeter (ECAL) targets electrons, photons, and some jets by measuring electromagnetic showers, whereas the Hadronic Calorimeter (HCAL), located outside the ECAL, is optimized for hadronic shower measurements.

The ECAL, depicted in its barrel section in Figure 2.11, consists of alternating layers of lead (Pb) and Liquid Argon (LAr). Its operation relies fundamentally on two physics processes: pair production and bremsstrahlung. These processes enable the calorimeter to accurately measure the energy and position of electromagnetic showers produced by particles such as electrons and photons.

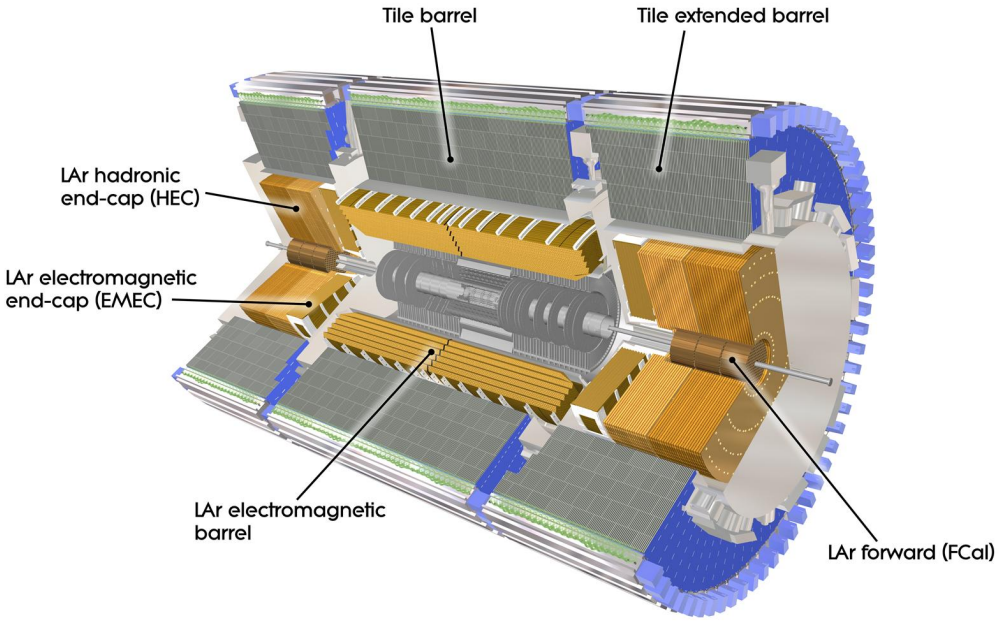


Figure 2.10: Scheme of the ATLAS calorimeter system [9]

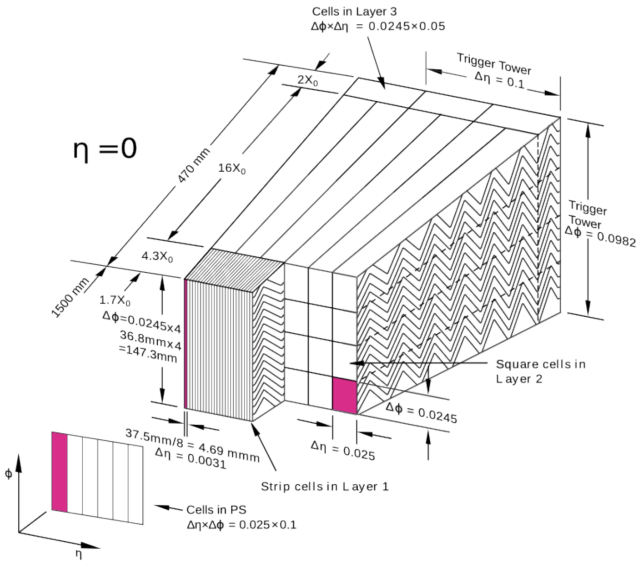


Figure 2.11: Scheme of the barrel module of the Electromagnetic Calorimeter in ATLAS with LAr active material and Pb as absorber [9].

High-energy photons traversing the calorimeter can undergo pair production, splitting into an electron-positron pair. Electrons and positrons emit bremsstrahlung radiation (high-energy photons) when deflected by nuclei's Coulomb fields. These photons can induce further electron-positron pairs, propagating the electromagnetic shower, which enhances the total energy deposition measured by the calorimeter.

The characteristic lengths associated with these processes are the radiation length X_0, which describes the mean distance over which an energetic electron's energy diminishes by a factor of $1/e$, and the pair-production length, roughly $9X_0/7$. These lengths depend on the medium's material properties, calculable via the empirical formula:

$$X_0 = \frac{1433[g\ cm^{-1}]\ A}{Z(Z+1)\ ln(183/Z^{1/3})}. \tag{2.10}$$

Divided in the barrel and end-caps components, covering respectively $|\eta| < 1.475$ and $1.375 < |\eta| < 3.2$, and each composed of LAr ($X_0 = 140\ mm$) and Pb ($X_0 = 5.6\ mm$) alternate layers, ECAL results in a total thickness which exceeds of 22 X_0 in the barrel and 24 X_0 in the end-cap regions, ensuring sufficient containment of electromagnetic showers.

Behind the ECAL lies the HCAL, which measures the energy of hadronic showers generated by strongly interacting particles, like pions. When neutral pions decay into two photons, these photons create electromagnetic showers within the calorimeter. Unlike electromagnetic interactions, hadronic showers involve complex nuclear interactions, secondary particle production, and nuclear fragmentation.

The parameter analogous to X_0 for hadronic interactions is the interaction length λ_I, representing the average distance a hadron travels before interacting with a nucleus. The ATLAS HCAL is subdivided into three regions: the Tile Calorimeter (TileCal) in the barrel, which uses steel as absorber and scintillating plastic tiles as active material; the Hadronic Endcap Calorimeters (HEC), employing LAr and copper; and the Forward Calorimeter (FCal), which incorporates copper and tungsten in the passive area. The combined depth of the HCAL is approximately 10 λ_I, providing full containment of particle showers along with the ECAL.

Energy resolution in calorimeters is characterized by the relation:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c\ , \tag{2.11}$$

where the terms symbolize statistical fluctuations (Poisson, shower sampling), electronic noise, and instrumental effects, respectively.

In the electromagnetic calorimeter, the energy resolution is roughly:

$$\frac{\sigma_E}{E} \simeq 10\%\sqrt{E} + 0.7\%, \tag{2.12}$$

which results poorer due to the showers complexity for the hadronic one with an approximated value of

$$\frac{\sigma_E}{E} \simeq 50\%\sqrt{E} + 3\%. \tag{2.13}$$

2.2.3 Muon Spectrometer

Since muons are heavier than electrons and interact only weakly with matter, they are among the few particles (excluding neutrinos) capable of passing through the inner detector and multiple layers of calorimeters with minimal energy loss. Additionally, their relatively long lifetime –attributable to their small mass– and their higher critical energy in various media mean that muons do not primarily lose energy through Bremsstrahlung radiation at LHC. Instead, they

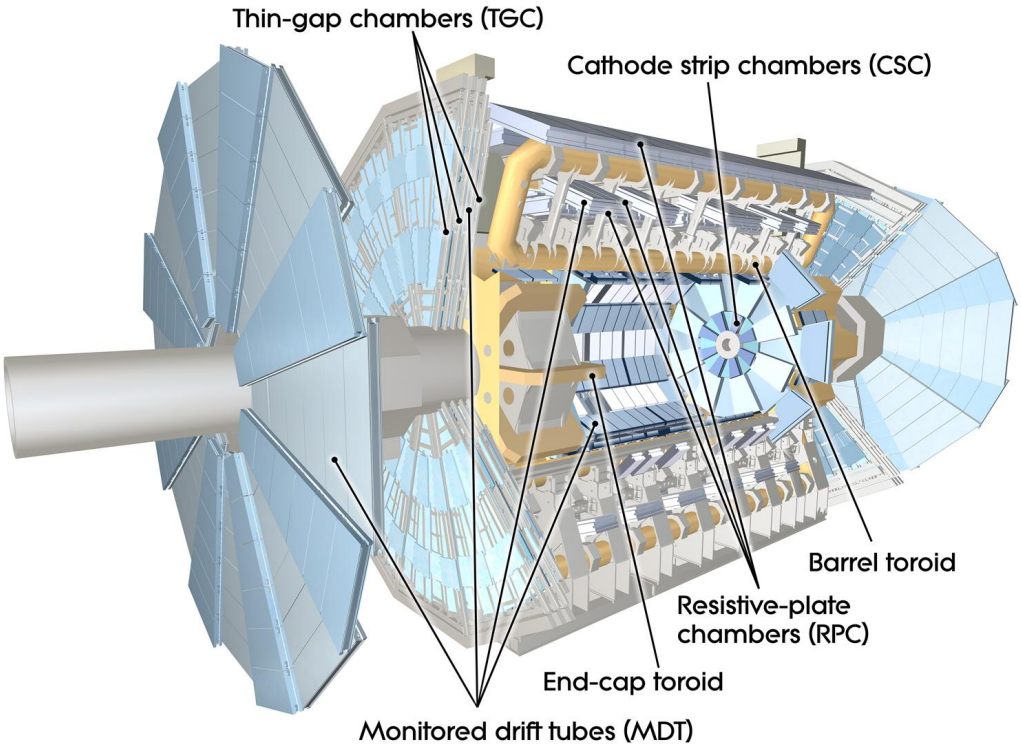


Figure 2.12: Muon spectrometer with magnet system scheme

behave as Minimum Ionizing Particles (MIPs), depositing a nearly constant amount of energy as they traverse material.

Based on these properties, the ATLAS Muon Spectrometer (MS) is positioned as the outermost sub-detector of the experiment. Its main purpose is to measure muon momentum by analyzing the curvature of their trajectories within the magnetic fields generated by specialized magnet systems across different layers.

The muon spectrometer is divided into two main regions: the barrel and the end-caps. The barrel region, covering the pseudorapidity range $|\eta| < 1.4$, uses the magnetic field generated by a large barrel toroid to bend charged particle trajectories. The end-cap regions, spanning $1.6 < |\eta| < 2.7$, employ two smaller end-cap magnets located at the ends of the barrel toroid to curve the muon tracks. The transition region, between $1.4 < |\eta| < 1.6$, utilizes both magnetic field systems. This setup ensures that the tracks are bent mainly in the R–z plane, thereby minimizing resolution degradation caused by multiple scattering as muons pass through the muon spectrometer layers.

The muon system, as depicted in Figure 2.12, employs various technologies for precision tracking and fast triggering. These include:

- Monitored Drift Tubes (MDT)
- Cathode Strip Chambers (CSC)
- Resistive Plate Chambers (RPC)
- Thin Gap Chambers (TGC)

The first two are high-resolution tracking detectors, while the latter are fast timing chambers optimized for trigger purposes.

The barrel region is organized into three stations: the Inner Barrel (around 5 meters from the beam axis), the Medium Barrel (about 7.5 meters), and the Outer Barrel (approximately

10 meters). MDT detectors are installed in all three stations, while RPCs are present only in the Medium and Outer Barrel layers. The end-cap regions are arranged on three separate wheels orthogonal to the beam axis, designed to provide high spatial resolution and efficient trigger performance.

Muon Spectrometer chambers for precision tracking

The precision tracking system comprises the Monitored Drift Tubes (MDT), which deliver accurate measurements across most of the η-range, and Cathode Strip Chambers (CSC), located in the innermost layer due to their higher rate capability and superior time resolution, which are essential in areas with the highest particle flux and track density.

MDTs are aluminum tubes approximately 3 cm in diameter, with lengths ranging from 1 to 6 meters. They are filled with an Argon-Carbon Dioxide ($Ar-CO_2$) gas mixture in a 93:7 ratio, containing a tungsten anode wire at their center. When a charged particle passes through, it ionizes the gas, producing electron-ion pairs that drift toward the anode wire and cathode. By measuring the drift time of these electrons, the impact point of the original particle can be reconstructed. However, because the signal primarily results from ion migration –taking up to 700 ns – these chambers are not suitable for first-level trigger purposes and require dedicated online systems for event selection. In the forward region, due to the low MDT's rate, CSCs detectors are used, as they works at higher rates compatible with the one expected in the end-cap region at the nominal luminosity and small angles. These detectors are no more installed in ATLAS, since they were implemented only in one of the end-cap wheels, the Small one, which now is replaced by the New Small Wheel. However, the data used in the analysis here presented were collected during the Run-2 phase, when CSCs were still working.

Muon Spectrometer chambers for fast trigger

The muon trigger system incorporates both the RPC, places in three layers in the barrel regions, and the TGC, organized in 3-4 layers in the end-cap ones. The RPCs consist of two resistive plates separated by a 2 mm gap, with an electric field of about 4.9 kV/mm, which facilitates the formation of charged avalanches along the muon's track. These avalanches are collected on the anode, allowing precise determination of both the η and ϕ coordinates. Thanks to the small gas gap and fast signal formation from electrons deposited on the strips, RPCs achieve a timing resolution of approximately 1.5 ns, making them ideal for the trigger system's needs. Their rapid response also enables efficient bunch crossing identification.

In the end-cap regions, the trigger system utilizes TGC detectors, which are multi-wire proportional chambers known for their excellent timing resolution. These chambers are primarily used for bunch crossing identification and also provide accurate measurements of the muons' position and momentum.

2.2.4 Trigger and data acquisition systems

At the LHC, proton bunches cross at the center of the ATLAS detector every 25 ns, corresponding to a nominal bunch-crossing rate of 40 MHz. Each of these crossings can give rise to multiple proton–proton interactions, especially at the high instantaneous luminosities reached during Run-2 and Run-3. As a result, billions of particles are produced every second, leaving signals across the tracking detectors, calorimeters, and muon systems.

The extremely high rate and complexity of the raw data produced by the ATLAS subdetectors make it both technically unfeasible and scientifically unnecessary to record every collision in full. During Run-2, storing the complete detector information for each event would have resulted in a data throughput of approximately 60 TB per second.

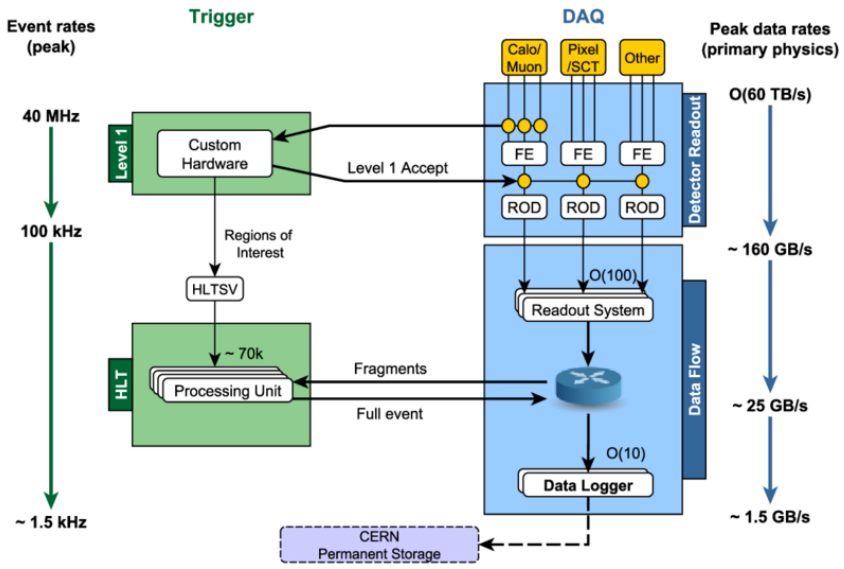


Figure 2.13: Trigger and Data Acquisition system used by ATLAS during Run-2 with the expected peak rates and bandwidths decrement through each component [12]

To address this challenge, ATLAS employs a dedicated Trigger and Data Acquisition (TDAQ) system, designed to rapidly identify potentially interesting collisions in real time and to reduce the raw event rate by several orders of magnitude, while retaining sensitivity to the widest possible range of physics signatures. The TDAQ system performs this selection hierarchically, first using custom hardware logic with a few-microsecond latency, and then with software-based algorithms running on large computing farms. This architecture reduces the initial $40MHz$ bunch-crossing rate to a final output of about 1–$2kHz$ of events stored permanently for offline reconstruction and analysis.

Precisely, the ATLAS trigger system during the Run-2, in which the data used in this thesis were taken, is composed as follows:

- **Level 1 (L1) Trigger**, located in the ATLAS detector is hardware-based and takes decisions based on raw inputs from a subset of information coming from calorimeters and muon spectrometer looking for simultaneous characteristic signals with less than $2.5\mu s$, keeping in the meanwhile the event information in buffer storage and defining preliminary regions of interest (RoI) in the detector;

- **High Level Trigger (HLT)**, a software-based trigger operating on dedicated computer farms which takes decisions using the entire detector information together with the specific RoI defined by the L1, permanently storing then the selected events and transmitting them to the CERN Tier-0 computing site for offline reconstruction.

With the L1 trigger the information rate is reduced from 40 MHz to 100 kHz, selecting about 100 000 events per seconds. The HLT selection further decrease the rate to about 1.5 kHz, for a final selection of about 1500 events per second which are then stored offline and used by the collaboration in physics analysis. A functional diagram of the ATLAS Trigger and Data Acquisition system is reported in Figure 2.13.

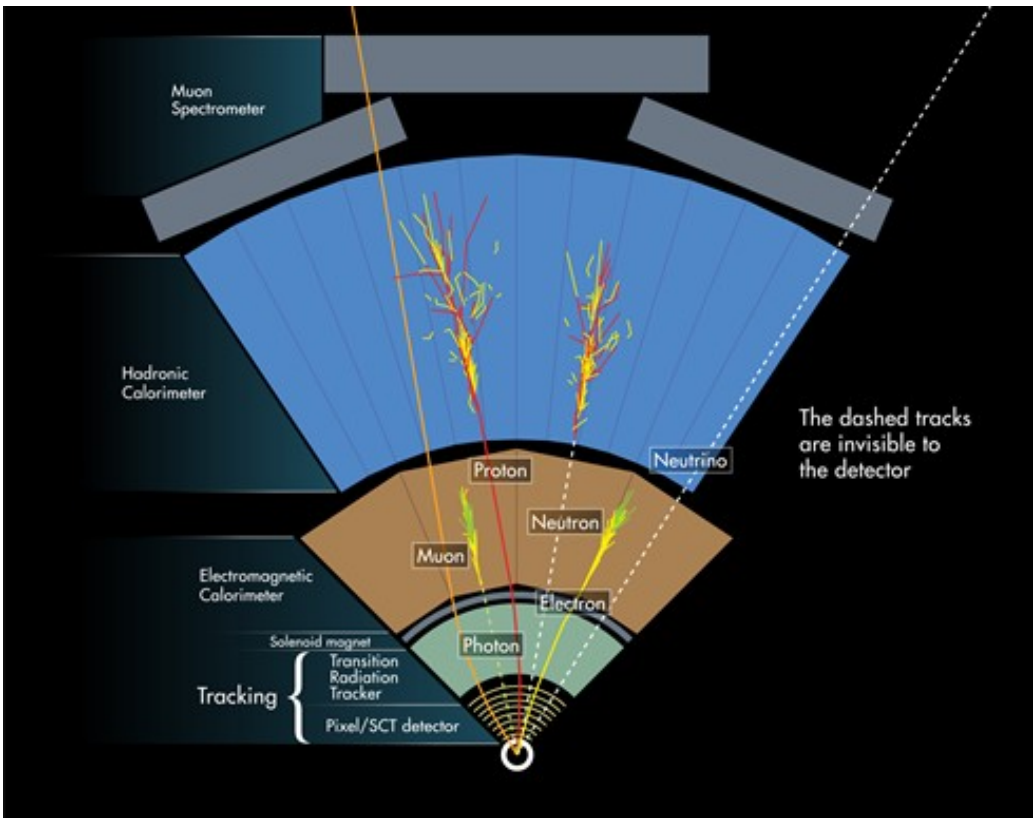


Figure 2.14: How different particles produce signatures within the ATLAS detector

2.3 Physics Objects reconstruction in ATLAS

As discussed in the previous section, particles produced in proton–proton collisions at ATLAS traverse the detector and leave characteristic signatures in the various sub-detectors. These signatures, illustrated in Figure 2.14, include hits on successive layers of the Inner Detector (ID) and the Muon Spectrometer (MS), as well as energy deposits in the electromagnetic and hadronic calorimeters.

From these raw responses, a first stage of reconstruction produces so-called *detector-level* objects, such as hits, clusters and space-points. These are then combined and processed to build higher-level *physics objects*, corresponding to individual particles (e.g. electrons, muons, photons, jets) that can be directly exploited in physics analyses.

The algorithms used for this reconstruction are centrally developed and maintained within the ATLAS software framework, `Athena` [13]. To ensure accuracy, reconstructed physics objects are subsequently calibrated, correcting for detector effects and residual discrepancies between simulation and data. This calibration step is essential to guarantee that the reconstructed quantities are suitable for precision measurements and searches for new physics.

In the following, the successive stages of this reconstruction chain will be presented in detail, starting from the definition of low-level detector signals and progressing to the formation of high-level objects such as leptons, photons, jets, and missing transverse momentum.

2.3.1 Tracking and Vertexing

The reconstruction of charged particles in ATLAS begins with the measurement of their passage through the layers of the ID. As a charged particle traverses the silicon sensors, it produces localized ionization, which is collected by the detector electronics and, if the signal is over the calibrated threshold, it is recorded as a **hit**. Each hit represents the most elementary piece of

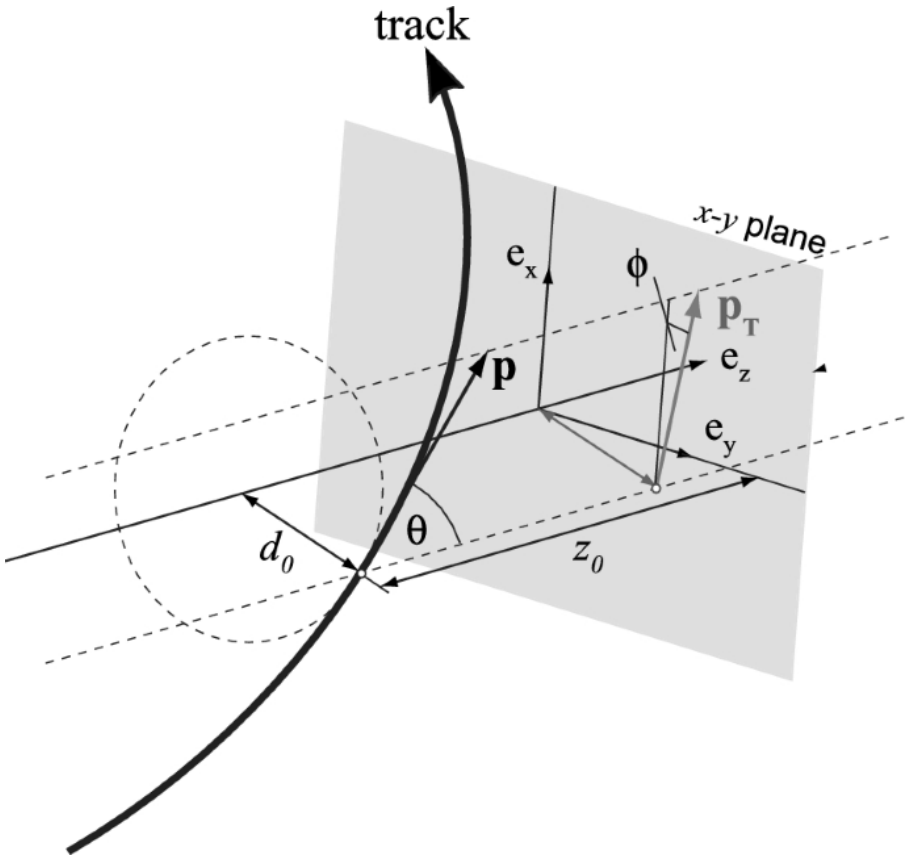


Figure 2.15: Schematization of the track parameters used by ATLAS [8]

tracking information and is associated with the precise position of the sensor element in which the energy deposition occurred in (η, ϕ) coordinates.

Because a particle deposit charge across multiple adjacent readout channels, the signals from neighboring channels are grouped together into **clusters**. Clustering improves the spatial resolution of the measurement by exploiting the distribution of the deposited charge, and it reduces the sensitivity to electronic noise. The position of the cluster is determined by a centroid calculation or more refined algorithms that take into account the profile of the collected charges in the pixel sensors.

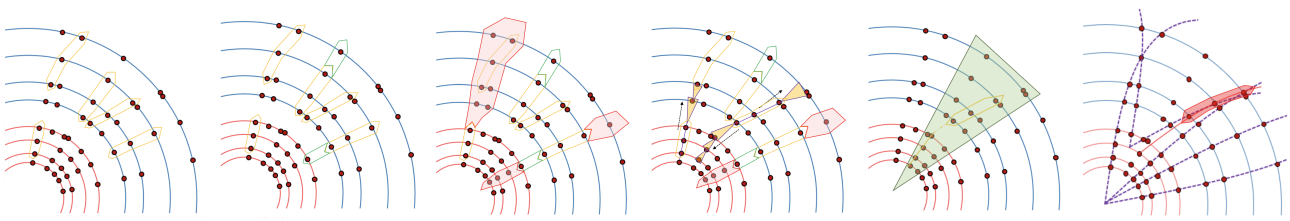
From each cluster the **space-point**, representing the point of intersection of the particle with the detectors, is then computed through a charge interpolation technique based on the charge distribution in different hits. Clusters from different layers are then combined into space-points, which provide a three-dimensional representation of the particle's trajectory at the location of the measurement. In the pixel detector, each cluster directly provides a space-point due to the two-dimensional segmentation of the pixels. In the silicon strip detector, clusters from two layers with different stereo angles must be matched to reconstruct a space-point in three dimensions.

The ensemble of space-points from multiple detector layers forms the input to the track reconstruction algorithms.

Tracks and vertices reconstruction in the ID

The helicoidal trajectory followed by the charged particles within the ID due to the uniform magnetic field produced by the solenoid are described, as reported in Figure 2.15, by five parameters: $(d_0,\ z_0,\ \theta,\ \phi,\ q/p_T)$.

d_0 and z_0 represent respectively the transverse and longitudinal distance of the interaction



(a) (b) (c) (d) (e) (f)

Figure 2.16: ATLAS *inside-out* track finding steps: (a) track seed formation from three space-points in the silicon detectors. The triplet provides an initial estimate of the trajectory parameters used for track finding; (b) seed confirmation through the addition of a fourth space-point that, when combined with the triplet, yields a compatible trajectory (for *fake-seeds* suppression); (c) *search roads* definition through subsequent detector layers. Only modules intersecting the predicted seed trajectory are considered, reducing combinatorics; (d) track extension using Kalman filter, where compatible clusters from adjacent layers are added to refine the trajectory, with branching when multiple hits are found; (e) bremsstrahlung recovery for electron tracks which allows to change track trajectory looking for seed within the calorimeters RoI and applying a Gaussian Sum filter; (f) final collection of potential track candidates obtained as output of the track finding stage

point, i.e. the point of the closest approach of the curve to the z-axis; ϕ and θ are the azimuthal and polar angle of the track direction at its perigee. The q/p_T quantity, which represents the charge of the reconstructed track divided by the transverse momentum of the track originating particle, is connected to the particle curvature due to the solenoidal magnetic field through the Equation 2.9.

The principal method employed in ATLAS for reconstructing tracks originating near the primary interaction point is the *inside-out* tracking algorithm. This approach exploits the fine granularity and excellent spatial resolution of the innermost silicon detectors to seed tracks, which are then progressively extended outward through the remainder of the tracking system.

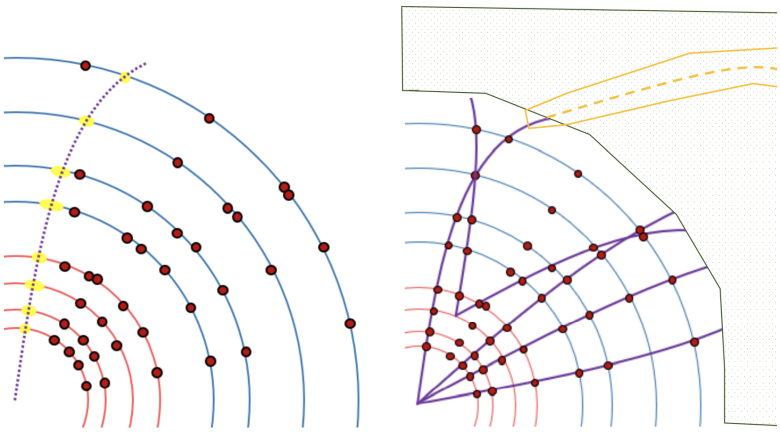
Seeding

Track finding begins with the formation of track seeds in the Pixel and SCT detectors. Seeds are constructed from combinations of two or three spatial points (3D hit positions), typically using triplets of hits when available, or doublets supplemented by constraints from the measured beam spot. The beam spot constraint significantly reduces the combinatorial background by rejecting seed candidates incompatible with the luminous region of the LHC. At this stage, an initial estimate of the track parameters (transverse momentum, direction, and impact parameters) is obtained with relatively large uncertainties.

Track Building

The seeds are then extended outward through successive silicon layers using a combinatorial Kalman filter [14]. At each detector layer, the predicted trajectory is extrapolated and compatible hits are sought within a search window defined by the propagated uncertainties. If multiple hits are consistent with the prediction, the algorithm branches, creating multiple track candidates from the same seed. Each candidate accumulates a quality score based on the number of hits, the number of missing hits ("holes"), and the fraction of shared hits with other candidates. Early rejection criteria are applied to suppress candidates with poor fit quality or excessive missing hits, thereby controlling combinatorial growth in dense environments such as high pile-up.

Additionally, potential seeds that fail the track finding are checked for calorimeter compatibility: if the seed is within a region of interest (RoI) in the calorimeter the track-finding



(a) (b)

Figure 2.17: ATLAS *inside − out* ambiguity solving and TRT extension steps: (a) diagram of the global χ^2 fitter step used for assigning a fit-related score to each candidate; (b) extension of silicon track candidates into the TRT through the Kalman filter based road search which add hits that improve the fit

procedure is redone allowing an additional "kink" in the track. This is known as Bremsstrahlung recovery and helps recover some efficiency for electron reconstruction.

TRT Tracks Extension

Surviving track candidates are propagated into the TRT. Although the straw tubes provide a coarser spatial resolution than the silicon detectors, they deliver a large number of hits per track (typically 20–30), which improves the overall momentum resolution and extends the tracking coverage to larger radii. The association of TRT hits is performed with relatively loose criteria to maximize efficiency; candidates for which TRT extension fails are retained as silicon-only tracks.

Ambiguity Resolution

The combinatorial nature of the pattern recognition produces multiple overlapping or duplicate track candidates. To resolve these ambiguities, a global scoring algorithm is employed. The score takes into account the number and type of associated hits (with higher weight for silicon hits), the fit χ^2, and penalties for missing or shared hits. Candidates are ranked by score, and conflicts are resolved by removing tracks that share excessive hits with higher-quality candidates. This step yields a unique and consistent set of reconstructed tracks.

Precision Track Fit

The final stage is a precision refit of the surviving track candidates, providing the optimal estimates of the track parameters.

Complementary Algorithms

The *inside-out* strategy is optimized for prompt tracks emerging from the primary vertex. To recover tracks that are not efficiently reconstructed in this framework, additional algorithms are employed. In particular, the *outside-in* tracking procedure begins from track segments in the TRT or muon spectrometer and propagates them inward, which is especially effective for displaced tracks and late photon conversions. Dedicated algorithms are also implemented for secondary vertices (e.g., $K^0_S \to \pi^+\pi^-$, $\Lambda \to p\pi^-$) and for very low-p_T particles whose trajectories curl significantly in the magnetic field.

Tracks used for physics analyses are required to fall within the ID acceptance ($|\eta| < 2.5$)

with a p_T greater that 0.5 GeV and have to and pass additional quality criteria such as the number of silicon detector hits and the association with the primary vertex.

Vertices are defined as the detector-level objects representing locations where interactions occur and are reconstructed targeting points within the detector from which multiple tracks originate.

The primary vertex (PV) is essentially the point where the primary proton-proton interaction occurs per bunch crossing, and is reconstructed through an iterative process of vertex finding and fitting [15] [16] in a procedure guided by the tracks impact parameters (d_0, z_0) which iteratively compute the compatibility of candidate tracks based on the vertex position. All the compatible tracks are then fitted again under the assumption that they originates from the computed vertex, and incompatible tracks are used as seeds for another vertex reconstruction, ending the procedure when no more candidate tracks are available for constructing new vertices.

Vertices are accepted if they are associated with at least two candidate tracks ($p_T > 0.5\ GeV$), and the primary one is usually identified as the one with the maximal $\sum_{i=0}^{Ntracks} p_{T_i}^2$.

2.3.2 Topological Calorimeter Clusters

While tracking and vertexing techniques excel at reconstructing the trajectories of charged particles and pinpointing their interaction vertices, a complete picture of the event requires the detection and measurement of neutral particles and the overall energy flow. The ATLAS calorimeter system addresses this need by capturing the energy deposited by particles that do not produce tracks, such as photons and neutral hadrons.

Topological clusters, which are defined as collections of adjacent calorimeter cells grouped based on their spatial proximity and the significance of their energy deposits, are essential in this context as they serve as the primary objects for calorimeter-based energy reconstruction. By grouping these cells, topological clusters enable precise 3D localization and measurement of particle showers within the active calorimeter volume. These clusters form the building blocks for reconstructing higher-level physics objects like jets, photons, and missing transverse energy, integrating calorimeter information with the tracking system to achieve comprehensive event reconstruction.

2.3.3 Electrons

Electrons are fundamental charged leptons that play a vital role in many physics analyses performed with the ATLAS detector. Their accurate reconstruction is essential for precise measurements of Standard Model processes and searches for new phenomena. The electron reconstruction process in ATLAS leverages the complementary information from the inner tracking detector and the electromagnetic calorimeter, enabling both identification and energy measurement with high efficiency and purity.

As shown in Figure 2.18, an electron originating from the interaction point first creates hits as it travels through the inner detector, with its trajectory curved by the magnetic field. Next, it enters the calorimeters, where it triggers an electromagnetic shower made up of secondary electrons and photons. This cascade continues until the electron's energy is entirely absorbed by the calorimeter material. The resulting shower exhibits a distinctive spatial and energy distribution pattern within the calorimeter, which can be exploited to identify the particle as an electron and to accurately determine its energy.

Electron reconstruction in ATLAS combines information from both the inner detector and the electromagnetic calorimeter. The process begins with the identification of tracks in the inner detector, reconstructed using pattern recognition algorithms that process hits in the pixel,

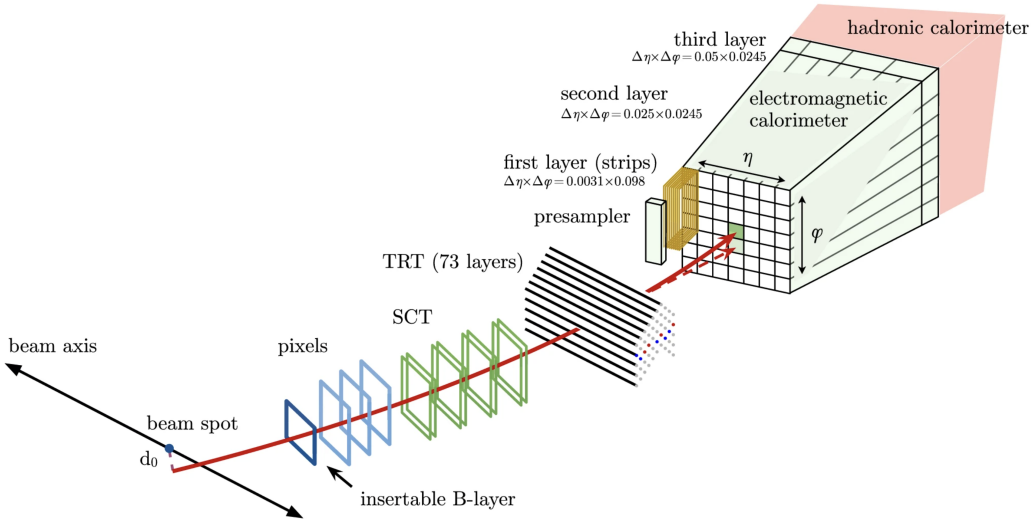


Figure 2.18: Schematic representation of the path of an electron through the detector. Red trajectory shows the hypothetical path of an electron, which first traverses the tracking system and then enters the electromagnetic calorimeter. Dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material in the tracking system [17].

silicon microstrip, and transition radiation tracker (TRT) layers. These tracks are extrapolated toward the calorimeter, where they are matched to energy deposits reconstructed as topological clusters based on calorimeter cells. The matching involves analyzing variables such as the angular separation between the track and the calorimeter cluster, the ratio of calorimeter energy to track momentum (E/p), and shower shape variables characteristic of electromagnetic interactions.

Once a track is matched to a calorimeter cluster, it forms an electron candidate. The subsequent step involves applying quality criteria to differentiate genuine electrons from fake ones. This discrimination is based on variables related to the electron's measurements in the inner detector and calorimeters, such as the impact parameter (d_0) and its significance ($|d_0/\sigma(d_0)|$), the number of hits associated with the track within the inner detector, and various shower shape variables that characterize the energy deposition.

Using these discriminating variables as inputs, a likelihood is constructed from their probability density functions. This likelihood yields a discriminant value, which can then be subjected to fixed cut thresholds to define different working points corresponding to varying electron identification efficiencies.

In physics analyses, it is important to differentiate between prompt and non-prompt electrons. Prompt electrons originate directly from the primary interaction vertex, typically produced in electroweak processes such as W or Z boson decays. Non-prompt electrons, on the other hand, result from secondary processes, such as decays of heavy-flavor hadrons (e.g., b or c quarks) or photon conversions occurring away from the primary vertex. Identifying whether an electron is prompt or non-prompt relies on various techniques, including impact parameter measurements, vertex association, and isolation criteria, which help suppress background and improve the purity of signals originating from the primary collision.

This comprehensive approach to electron reconstruction ensures high efficiency and purity, providing a reliable foundation for detailed physics measurements within the ATLAS experiment.

After the initial reconstruction and identification, the measured electron energy in the calorimeter requires calibration to ensure accurate and precise energy estimates. This calibration process corrects for various detector effects, such as non-compensating calorimeter

response, energy losses in inactive material, and variations in detector response across different regions.

The calibration procedure involves a sequence of steps to adjust the measured energy to correspond to the true electron energy at the interaction point. Typically, this starts with a reference calibration derived from test-beam data and Monte Carlo simulations, which models the detector response. Further corrections are applied using in-situ techniques, such as comparisons of reconstructed electron energies with well-known physics processes like $Z \to e^+e^-$ decays, to account for residual discrepancies and regional variations.

The calibration factors are determined using sophisticated algorithms that minimize the differences between the reconstructed and true energies, often employing multivariate approaches or machine learning techniques. The goal is to produce a calibrated electron energy scale that is both linear and stable across the detector geometries and energy ranges, ultimately leading to improved measurement precision in physics analyses.

2.3.4 Photons

Unlike electrons, photons do not leave tracks in the inner detector because they carry no electric charge. However, once they enter the electromagnetic calorimeter, they initiate electromagnetic showers. Within the calorimeter, photons mainly interact via pair production and Compton scattering, which produce electron-positron pairs and scattered photons. These secondary electrons and positrons continue to interact with the calorimeter material, generating additional electrons and photons and thus sustaining the development of the shower. Ultimately, the entire energy of the initial photon is absorbed by the calorimeter, similarly to the energy deposition process observed for electrons.

Due to their similarities, photon reconstruction proceeds in parallel with the electron one. Photons can be split in two categories, the *converted* photons, which are the one that generate e^+e^- pairs within the inner tracker volume that will produce showers in the ECAL, and the *unconverted* photons, that do not undergo any conversion process before reaching the electromagnetic calorimeter, meaning they produce no associated track or only very minimal activity in the tracking system.

Converted photons can be identified by the detection of a conversion vertex originating from a pair of tracks with opposite charges (curvature), and reconstruction algorithms look for these track pairs originating from a common point, along with consistency checks with the calorimeter energy deposit. On the other hand, unconverted photons are identified mainly through their calorimetric shower shape characteristics and the absence of matching tracks. For these, the primary signature is a localized, symmetric electromagnetic shower in the calorimeter with no associated tracks in the inner detector.

An ambiguity solver is also employed to distinguish between electrons, converted photons and unconverted ones.

2.3.5 Muons

Owing to their penetrating nature, muons are uniquely capable of traversing the entire detector, reaching the outer muon spectrometer. This property is fundamental to their identification in ATLAS. Muon reconstruction leverages information from both the inner tracking system and the muon spectrometer, combining these measurements to achieve high efficiency and purity.

The process begins with the reconstruction of tracks in the inner detector, which provides precise measurements of a particle's initial trajectory and momentum close to the interaction point. In parallel, the muon spectrometer, situated outside the calorimeters, reconstructs muon

tracks based on signals from its various chambers, including drift tubes, resistive plate chambers, and calorimeter trigger chambers.

Muon track reconstruction in the MS employs a multi-stage approach. Initially, pattern recognition algorithms process the raw hit data to identify linear segments that could originate from muons traveling through the chambers. A common technique used is the Hough transform, a mathematical method that maps the spatial hit information into a parameter space to identify lines or curves corresponding to potential muon trajectories. This method effectively filters out noise and spurious hits, reliably extracting candidate track segments even in high-occupancy scenarios.

Once these segments are identified, further fitting procedures refine the track parameters, including the muon's position, direction, and momentum. Multiple segments from different chambers are then combined to form a global muon track, which is subsequently parameterized using techniques such as Kalman filtering. These compiled tracks are crucial for precise momentum measurements and serve as the primary data for muon identification.

The next step involves associating the tracks reconstructed in the inner detector with those from the muon spectrometer. This matching process compares track parameters –such as momentum, direction, and spatial extrapolations– using criteria like the spatial separation of the trajectories, the consistency of the momentum measurements, and vertex compatibility. Successful matches produce high-quality muon candidates, whose combined track fit yields the best estimate of the muon's kinematic properties.

Although muons are minimally ionizing particles, depositing only a small amount of energy in the calorimeters, these signatures are valuable for background rejection. Low calorimeter energy deposits aligned with the muon candidate provide confirmation of the particle's muonic nature and help discriminate against hadronic backgrounds.

Additional quality criteria (based on the number of hits in each detector component, fit quality metrics, and the consistency between the inner detector and muon spectrometer measurements) are applied to refine the selection further. Advanced classification techniques, including likelihood-based and multivariate approaches, are employed to optimize the balance between efficiency and background suppression.

This integrated reconstruction strategy, leveraging sophisticated algorithms like the Hough transform for MS segments, comprehensive track fitting, and multimodal detector information, ensures a clean, efficient, and well-understood muon sample. This robustness is essential for attaining the precise measurements and background control needed in the diverse analyses conducted by ATLAS.

2.3.6 Jets

Because of color confinement, free color charges cannot be directly observed. Instead, high-energy quarks and gluons undergo hadronization, transforming into a shower of mainly hadrons, but also photons and leptons. This process is reconstructed as a jet, which is a narrow, collimated spray of particles resulting from the hadronization of a high-energy quark or gluon produced in proton-proton collisions.

A precise reconstruction of jets is essential for nearly all physics analyses at hadron colliders, and it is particularly important in Higgs boson studies. Many Higgs decay channels involve hadronic final states. In these cases, accurately measuring the jet kinematics is critical to reconstruct the invariant mass of the Higgs candidate, suppress backgrounds, and achieve optimal signal sensitivity.

The general concept behind jet reconstruction is to cluster the particles produced in hadronization in such a way that they reflect their common mother parton. By grouping final-state particles that originated from the same high-energy quark or gluon, jet algorithms aim to ap-

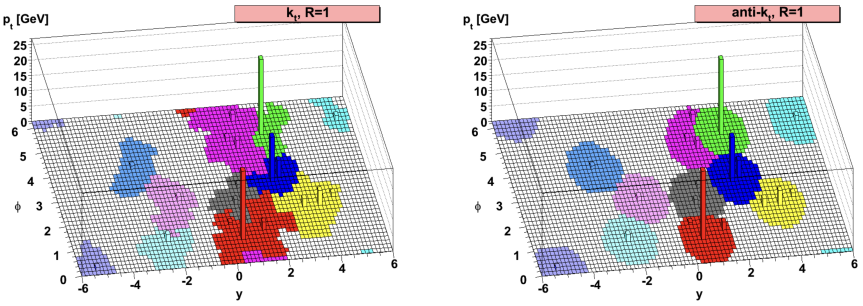


Figure 2.19: Reconstruction in the $y - \phi$ plane of a sample parton-level event with the (left) k_t algorithm and the (right) anti-k_t algorithm with $R = 1$ parameter [18].

proximate the kinematics of the underlying parton, providing a bridge between the observed detector-level signatures and the fundamental interactions described by QCD. A robust jet algorithm must also be resilient to detector effects, soft radiation, and pile-up, while preserving theoretical properties such as infrared and collinear safety, enabling meaningful comparisons with perturbative QCD predictions.

To achieve this goal, distance-based algorithms were developed to cluster the hadronization products into jets using their measured kinematic properties (p_T, η, ϕ). These algorithms rely on the definition of two fundamental distances: the distance between two particles, $d_{i,j}$, and the distance between a particle and the beam, $d_{i,B}$.

$$d_{i,j} = min(p_{T_i}^{2P}, p_{T_j}^{2P}) \frac{\Delta R_{i,j}^2}{R} \quad , \quad d_{i,B} = p_{T_i}^{2P}. \tag{2.14}$$

The distance between particles is typically defined using $\Delta R^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where R is the radius parameter controlling the typical jet size (e.g. $R = 0.4$ or $R = 1.0$).

Jet clustering algorithms operate by identifying the smallest distance between two particles and then merging them into a single cluster. After each merging step, the algorithm recalculates the distances to include the newly formed cluster. This iterative process continues until all particles have been successfully grouped into jets or classified as individual jets.

The exponent P in Equations 2.14 determines the specific variant of the clustering algorithm: for $P = 0$, the algorithm depends purely on the geometric separation of particles; for $P = 1$, the k_t algorithm preferentially merges softer particles first; and for $P = -1$, the so-called anti-k_t algorithm merges the hardest particles first, producing more regular, cone-like jets.

In this work, we focus exclusively on the anti-k_t algorithm, as it is the standard choice for jet reconstruction in ATLAS analyses due to its robustness against soft radiation and pile-up, and its favorable experimental properties.

Anti-k_t Algorithm for jet clustering

Setting $p = -1$ the two distance measures become

$$d_{i,j} = min(p_{T_i}^{-2}, p_{T_j}^{-2}) \frac{\Delta R_{i,j}^2}{R} \quad , \quad d_{i,B} = p_{T_i}^{-2}. \tag{2.15}$$

This implies that particles with high transverse momentum have a smaller distance to the beam, d_{iB}, and therefore act as *attractors* for nearby softer particles, which tend to cluster around them. In practice, in the anti-k_t variant, once a hard particle is found, it progressively

absorbs all softer particles within a radius R. The resulting jets thus have a well-defined, nearly circular area in the (y, ϕ) plane, corresponding to a cone of radius R.

Having a well-defined and stable jet area is particularly advantageous, as it allows for a more effective pile-up mitigation: the contribution of additional soft interactions can be more easily estimated and subtracted when the jet boundaries are regular and known. Furthermore, the presence of a geometrically well-behaved jet facilitates a more consistent and reliable calibration procedure, since the jet response can be characterized with higher precision.

This is not always the case for other clustering algorithms such as the k_t algorithm. Because it merges soft particles first, the boundaries of k_t jets are often irregular, leading to jets with unstable shapes and variable areas. In contrast, the anti-k_t algorithm produces jets with smooth, conical shapes that are less sensitive to fluctuations in soft radiation.

In addition, the anti-k_t algorithm is both **infrared and collinear safe**. The addition of a soft particle does not modify the shape of an existing jet, since its low transverse momentum makes both d_{iB} and d_{ij} large; the particle is thus clustered into an existing jet without affecting its structure or creating spurious jets. Similarly, if a particle splits collinearly into two with nearly identical directions, their separation ΔR_{ij} is close to zero, making d_{ij} negligible. The two particles are therefore immediately re-merged, preserving the original jet configuration.

The regular and conical structure of the reconstructed jets, together with the stability and theoretical safety properties discussed above, makes the anti-k_t the standard jet reconstruction algorithm in the ATLAS experiment.

In ATLAS, the anti-k_t algorithm is implemented using the `FastJet` package [19], which provides an efficient computational framework for jet clustering. Depending on the chosen R value, one can distinguish in ATLAS between small radius (small-R) jets, with $R = 0.4$, and large radius (large-R) jets, with $R = 1.0$. The choice between small-R and large-R jets in a physics analysis primarily depends on the topology of the targeted events. Considering a particle X, characterized by its mass m_X and transverse momentum p_{T_X}, decaying into two daughter particles i and j, the angular separation between the decay products, ΔR_{ij}, can be approximated in the quasi-collinear limit by

$$\Delta R_{ij} \simeq \frac{1}{\sqrt{z(1-z)}} \frac{m_X}{p_{T_X}}, \tag{2.16}$$

where z denotes the transverse momentum fraction carried by each of the decay products, $z = p_{T_i}/p_{T_X}$.

In the specific case of a symmetric decay, such as $H \to b\bar{b}$, where the two daughter particles share approximately equal transverse momentum ($z \simeq 1/2$), the above expression simplifies to

$$\Delta R_{ij} \simeq \frac{2m_X}{p_{T_X}}. \tag{2.17}$$

This relation shows that the angular separation between the decay products decreases as the transverse momentum of the parent particle increases.

For example, in the $H \to b\bar{b}$ decay, one of the key processes studied in this thesis depicted in Figure 2.20, at low Higgs transverse momentum, referred as *resolved* topology, the two b-quarks are sufficiently separated to be reconstructed as two distinct small-R jets. However, in the so-called *boosted* topology, where the Higgs boson has a large transverse momentum, the decay products become highly collimated, making their reconstruction as two separate jets challenging. In this regime, the ATLAS collaboration employs a novel approach, described in more detail in Chapter 3, in which the Higgs boson decay is reconstructed using a single large-R jet that captures both b-quarks within the same large-R jet cone.

As in this thesis only boosted topologies will be considered, the only used jets are variable radius jets (VR) and large-R jets with $R = 1.0$.

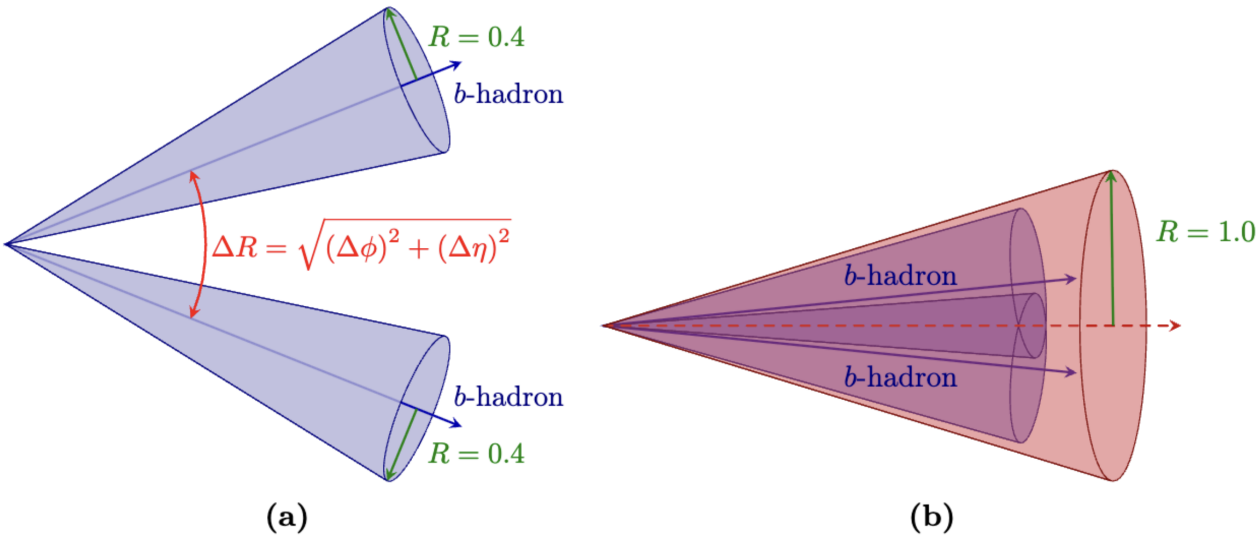


Figure 2.20: $H \to b\bar{b}$ decay: (a) small-R jets are used for reconstructing the event in the *resolved* topology; (b) in the *boosted* topology a single large-R jet is used for reconstructing the event [20].

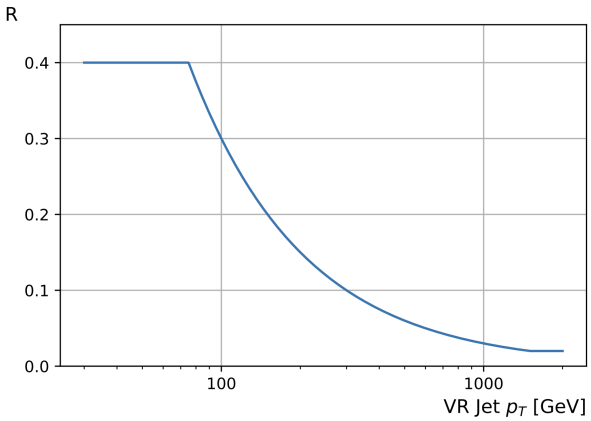


Figure 2.21: anti-k_t radius of VR track-jets as a function of the jet p_T [22].

Variable Radius jets

Variable Radius (VR) track-jets, described in Ref. [21], are reconstructed from inner detector tracks using a modified anti-k_t algorithm in which the effective jet radius depends on the transverse momentum of the jet:

$$R = \frac{\rho}{p_T}, \tag{2.18}$$

where the parameter $\rho = 30GeV$ is chosen to optimize the angular matching between the reconstructed subjet and the true b-hadron direction. The radius is constrained within $0.02 < R < 0.4$ to avoid pathological behavior at extreme p_T. This dynamic radius naturally decreases with increasing boost (as in Figure 2.21), ensuring that the two b-hadron decay products remain resolvable even when the Higgs boson is highly boosted.

The use of variable-radius subjets offers several key advantages over fixed-radius definitions. By adapting the jet size to the subjet transverse momentum, VR subjets maintain sensitivity to the separate b-hadron decay chains across a wide kinematic range, leading to improved Higgs-jet identification efficiency and reduced dependence on the jet boost. The adaptive radius also ensures that the jet captures all relevant decay products without including extra contamination.

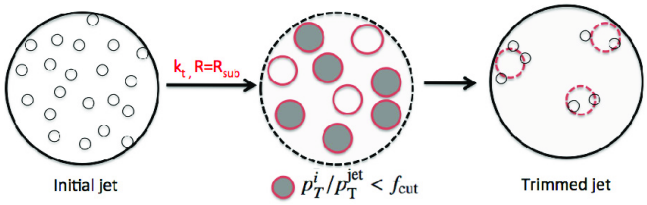


Figure 2.22: Diagram depicting the jet trimming procedure [23]

VR subjets therefore represent the Run-2 paradigm for subjet reconstruction within large-R jets and provide the foundation for flavour tagging.

UFO Large-R jets

To reconstruct the hadronic decays of highly boosted heavy particles such as the Higgs boson, the ATLAS experiment employs large-radius jets that encompass the full decay system. In Run-2 analyses, these jets were reconstructed using the anti-k_t algorithm with a fixed radius parameter of $R = 1.0$, taking topological calorimeter clusters as inputs. To suppress contributions from pile-up and the underlying event, the jets were groomed using a *trimming* algorithm. The trimming procedure, depicted in Figure 2.22, consist in a reclustering of the jet constituents into small radius subjets with $R_{\text{sub}} = 0.2$. Then subjets carrying less than 5% of the original jet transverse momentum are removed and the remaining ones are recombined into a trimmed jet. The resulting trimmed large-R jets retain the hard core of the boosted object while discarding soft, diffuse radiation.

In Run-3, the variable-radius approach evolved toward a unified reconstruction framework based on Unified Flow Object (UFO) jets, which combine tracking and calorimetric information at the constituent level. UFOs generalize the Particle Flow (PFlow) concept that was originally introduced during Run-2, providing a more flexible approach to jet reconstruction and enhancing robustness in high pile-up conditions. The PFlow algorithm reconstructs individual stable particles by combining measurements from the inner tracking detectors and the calorimeters, using the superior momentum resolution of tracks for charged particles and calorimeter information for neutral energy deposits. Charged-particle energy deposits are subtracted from the calorimeter to prevent double counting, and the resulting PFlow objects serve as inputs to jet reconstruction algorithms. This hybrid approach improves jet energy and angular resolution, enhances pile-up mitigation, and provides a more accurate modeling of jet substructure compared to purely calorimeter-based jets.

Building upon this concept, the UFO framework integrates information from all relevant subsystems, including tracking detectors and calorimeters, into a single, consistent representation of the event's final-state particles. This unified object model allows for the reconstruction of both small- and large-radius jets using the same set of calibrated inputs, thereby improving energy resolution, pile-up suppression, and the consistency of physics observables across jet collections.

UFO jets are reconstructed using the anti-k_t algorithm with a radius parameter of $R = 0.4$ and $R = 1.0$ respectively for small-R and large-R jets, using as input PFlow objects together with track calorimeter clusters, which are topological clusters in which also the matched track η and φ are taken into account. As in the Run-2 scheme, these jets are groomed using trimming [24], in which the constituents are reclustered into k_t subjets with radius $R_{\text{sub}} = 0.2$, and subjets carrying less than $f_{\text{cut}} = 5\%$ of the original jet p_T are removed. This grooming procedure suppresses pile-up and soft radiation while preserving the hard, collimated energy flow characteristic of boosted objects. The resulting trimmed UFO large-R jets therefore serve as the modern equivalent of the previous Local Cell Weighitn (LCW) topocluster- or PFlow-based

large-R jets.

The inclusion of charged-particle tracking information at the constituent level provides significant performance gains. The tracking system offers superior angular resolution and precise reconstruction of charged-particle momenta, while the calorimeter measures neutral energy deposits. By combining these complementary inputs, UFO jets achieve a more accurate and stable energy response, particularly in the high-pile-up conditions typical of Run-3. Additionally, per-object pile-up identification and machine-learning-based track–cluster linking improve the robustness of jet reconstruction and substructure observables.

UFO large-R jets are now used as the default inputs for boosted-object tagging in Run-3 analyses, including the boosted $GN2X$ tagger used for jet classification in the $VHb\bar{b}$ analysis described in Chapter 5. The UFO reconstruction thus provides a coherent and high-precision basis for both small- and large-R jet reconstruction, enabling consistent performance comparisons between the tagging algorithm which is calibrated and included into analysis in this work (which uses large-R jets) with the previous version of it (with small-R).

2.3.7 Missing Energy

The missing transverse momentum ($\vec{E}_T^{miss}$) is a key observable in ATLAS, sensitive to invisible particles such as neutrinos or possible dark-matter candidates. It is defined as the negative vector sum of the transverse momenta of all reconstructed and calibrated physics objects in the event, complemented by contributions from low-p_T activity not associated with these objects. Its magnitude is referred to as the missing transverse energy E_T^{miss}.

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\mathrm{objects}} \vec{p}_{\mathrm{T}}^{\,i} - \vec{p}_{\mathrm{T}}^{\,\mathrm{soft}}. \tag{2.19}$$

An object-based reconstruction approach is adopted, where electrons, muons, photons, hadronic τ candidates, and jets are included following a well-defined priority scheme to avoid double counting. The remaining contribution, referred to as the soft term, accounts for tracks and calorimeter deposits not associated with any high-p_T object. In Run-2 and Run-3, the soft term is typically built using the *Track Soft Term* (TST) algorithm, which relies on tracks originating from the primary vertex.

This method provides good stability against pile-up effects, as tracks from additional interactions are excluded. Alternative definitions, such as the Particle Flow (PFlow) missing transverse energy, combine tracking and calorimeter information at the particle level to improve resolution and reduce sensitivity to pile-up.

Chapter 3

Jet Flavour Tagging in ATLAS

The identification of the flavour of jets, commonly referred to as *flavour tagging*, is a key component of many physics analyses at the LHC. In particular, the ability to efficiently identify jets originating from the hadronization of bottom (b) or charm (c) quarks, while rejecting jets initiated by light quarks or gluons, is crucial for a broad range of measurements and searches. Examples include the reconstruction of top-quark decays, precision measurements of Higgs boson couplings, and searches for processes beyond the Standard Model that involve heavy-flavour signatures.

In the context of Higgs boson physics, as showed in Section 1.3, the dominant decay mode of the Higgs into a pair of b quarks provides a direct probe of the Yukawa coupling to down-type quarks. However, this final state suffer from an overwhelming QCD multijet background, making b-jet identification essential to isolate the signal.

Flavour tagging thus represents one of the most powerful tools to enhance the sensitivity of analyses such as the associated production of the Higgs boson with a vector boson or a top-quark pair.

Before delving into the algorithms designed to identify heavy-flavour jets, it is important to understand the fundamental signatures that distinguish b- and c-jets from light-flavour jets. These properties —such as long lifetimes, displaced decay vertices, and semileptonic decays— form the basis for all modern flavour-tagging strategies. Understanding these observables provides the necessary context for appreciating the evolution of tagging algorithms from classical vertex-based methods to modern graph neural networks optimized for identifying heavy-flavour jets across all the p_T spectrum.

3.1 Heavy-Flavour Signatures

The identification of heavy-flavour jets exploits the distinctive characteristics of hadrons containing bottom or charm quarks. These arise from their comparatively large masses and long lifetimes relative to hadrons formed from light quarks. Bottom hadrons have typical masses of about 5 GeV and lifetimes of approximately 1.5 ps, corresponding to a proper decay length of $c\tau \simeq 450$ μm. Charm hadrons are lighter (~ 2 GeV) and shorter-lived (~ 1 ps), leading to similar but less pronounced signatures.

The long lifetime of b-hadrons results in measurable displacements of their decay vertices from the primary interaction point. The average distance travelled before decay can be expressed as

$$d = \gamma\beta c\tau \simeq \gamma c\tau, \tag{3.1}$$

where $\gamma = E_b/m_b$ and $\beta = v/c \approx 1$ in the high-energy limit. For a b-hadron with an energy of 50 GeV, this corresponds to a mean decay length of approximately 4.5 mm, while a 1 TeV b-hadron would travel about 90 mm before decaying, well beyond the innermost pixel layer of

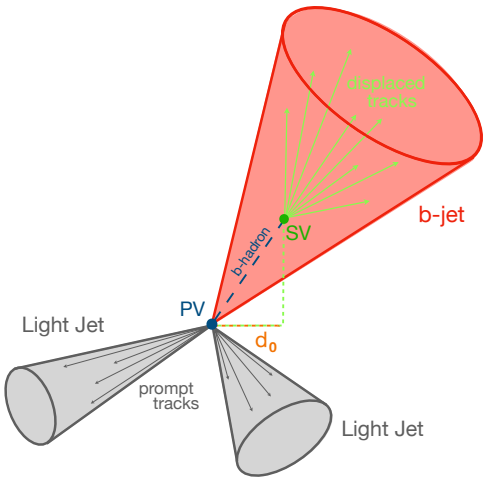


Figure 3.1: Diagram of a typical b-jet (red) produced in association with two light-jets (grey). The produced b-hadron (blue) has travelled a significant distance (dashed blue line) from the primary vertex (PV) before its decay in the secondary vertex (SV), where it produced the observed displaced tracks (light green). The large transverse impact parameter d_0 is one of the characteristic properties of b-hadrons decay products.

the ATLAS detector. Such sizeable displacements are a defining feature of b-jets and enable the reconstruction of secondary vertices (SV) at the b-hadron decay point.

These fundamental properties give rise to three main classes of observables that underpin modern flavour-tagging algorithms: **track-based impact-parameter measurements**, **secondary-vertex topologies**, and **semileptonic decay signatures**.

Secondary vertices The long lifetime and multi-body decays of b-hadrons lead to the formation of secondary vertices, which can be reconstructed by clustering displaced tracks that are geometrically and kinematically compatible with a common origin distinct from the primary vertex. The fragmentation of the b-quark into a b-hadron is described by the fragmentation function, which determines the fraction of the parent quark energy transferred to the hadron. This energy fraction directly impacts the kinematic properties of the b-hadron decay products and, consequently, the fraction of the jet energy carried by the tracks associated with the secondary vertex, shaping the characteristic topology of heavy-flavour decays. In many cases, the decay chain involves multiple vertices; for example, a b-hadron decaying into a charm hadron that subsequently decays further downstream, that can be reconstructed and used for distinguish heavy-flavour jets from the light-flavour ones.

Track Impact Parameters Charged-particle tracks reconstructed in the ATLAS Inner Detector are characterised by their transverse (d_0) and longitudinal (z_0) impact parameters, defined as the distances of closest approach of the track to the primary vertex in the transverse plane and along the beam axis, respectively. Tracks originating from the decays of long-lived b-hadron typically exhibit significantly larger impact parameters[1] than those produced directly at the primary vertex. Because b-hadrons often decay through multiple steps (for example $b \to c \to s$), several tracks in a b-jet can have large, correlated impact parameters. In contrast, tracks in light-flavour jets generally originate promptly, with impact-parameter significances

[1]In this discussion, the impact parameter refers to the transverse impact parameter d_0 and to the longitudinal impact parameter projected onto the transverse plane, $z_0 \sin\theta$. While d_0 directly probes the transverse displacement of the decay vertex, the use of $z_0 \sin\theta$ reduces the dependence on the track polar angle, providing a more robust measure of the longitudinal displacement.

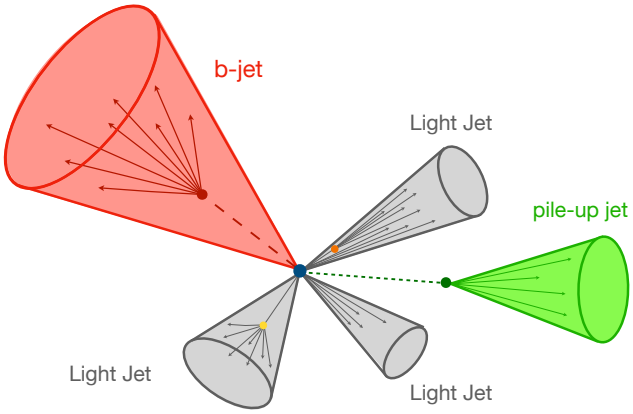


Figure 3.2: Illustration of processes that can mimic heavy-flavour signatures in light-flavour jets. Fake secondary vertices may arise from random combinations of tracks in dense environments (orange), while genuine secondary vertices can originate from interactions with the detector material (yellow), photon conversions, or decays of long-lived particles. Pile-up jets associated with additional primary vertices are shown in green.

distributed symmetrically around zero.

Semileptonic Decays Approximately 40% [25] of b-hadron decays are semileptonic, producing a soft muon or electron within the jet. Identifying these soft leptons provides an additional handle for flavour tagging, as they are rare in light-flavour jets.
c-jets share many of the same characteristics as b-jets, but with less pronounced features. Charm hadrons are lighter, have shorter lifetimes, and produce decays with lower track multiplicities. In addition, the c-quark fragmentation function is softer than that of beauty, resulting in charm hadrons carrying a smaller fraction of the jet energy and further reducing the displacement and kinematic signatures associated with secondary vertices.

As a consequence, the resulting jet topologies occupy an intermediate regime between those of b-jets and light-flavour jets, making the discrimination of c-jets challenging both with respect to beauty and light flavours.

Overall, distinguishing heavy-flavour jets from light-flavour jets can be challenging, as certain light-jet processes can mimic the signatures of heavy-flavour decays. As shown in Figure 3.2, displaced vertices may also appear in light-flavour jets due to several effects, such as random combinations of tracks, interactions of particles with the detector material, photon conversions, the presence of long-lived particles, or the erroneous association of tracks originating from additional pile-up collisions to the jet under study.

The following sections will initially describe the different tagging algorithms developed across the years. Then the strategies employed to achieve high performance flavour tagging in the resolved topology, emphasizing the techniques and algorithms, such as GN2, that are optimized for well-separated jets. Next, the focus will shift to the boosted topology, introducing ad-hoc strategies like the initial Xbb tagger, which ultimately leads to the GN2X tagger, highlighting its approach to leveraging jet substructure in dense, high-p_T environments. Its performance will be compared to alternative methods, including the use of variable-radius (VR) jets analyzed separately. Subsequently, a dedicated chapter will discuss the calibration techniques necessary to make GN2X suitable for physics analyses. Finally, in Chapter 5, we will underscore the advantages of a high-performance bb-boosted tagger in the $VH(bb)$ boosted analysis, illustrating how improved b-jet identification enhances the analysis of complex event

topologies.

3.2 Introduction to Flavour Tagging Strategies

The distinct signatures of heavy-flavour decays motivate the design of a variety of tagging algorithms, which in principle can divided in two macro categories: low-level and high-level algorithms. **Low-level algorithms** are the first step in the b-tagging process. Taking as input information about the jet and its associated tracks, these algorithms are a combination of manually optimized reconstruction algorithms which reconstruct the characteristic features of the identified jets such as the displaced decay vertices or the impact parameters. Examples include the SV1 [26], JetFitter [27] and the IPxD [28] algorithms. SV1 focused on reconstructing secondary vertices within jets to provide information about displaced decay points characteristic of b-hadron decays. JetFitter aimed to identify the topology of the b-hadron decay chain by reconstructing its flight path and related vertex features. IPxD combined impact parameter information from tracks to evaluate their likelihood of originating from secondary vertices.

Additionally, machine learning models such as RNNIP [29] and DIPS [30] were trained to predict jet flavour by using impact parameter and hit information from a variable number of tracks.

High-level algorithms, as depicted as example in Figure 3.3, integrated the outputs of these independently optimized low-level methods using multivariate techniques to produce a discriminant score for each jet. The DL1 family of taggers follows this paradigm, combining the outputs of multiple low-level flavour taggers using a neural network trained to distinguish between b-, c- and light-flavour jets.

The original DL1 algorithm relies exclusively on traditional, manually optimised low-level taggers such as IPxD, SV1, and JetFitter. The DL1r variant extends this approach by additionally incorporating the output of the trained track-based neural network RNNIP, enhancing the use of detailed track information. Finally, DL1d further reduces the dependence on handcrafted taggers by primarily using deep, trainable low-level models such as RNNIP and DIPS, marking a transition towards a more data-driven flavour-tagging strategy.

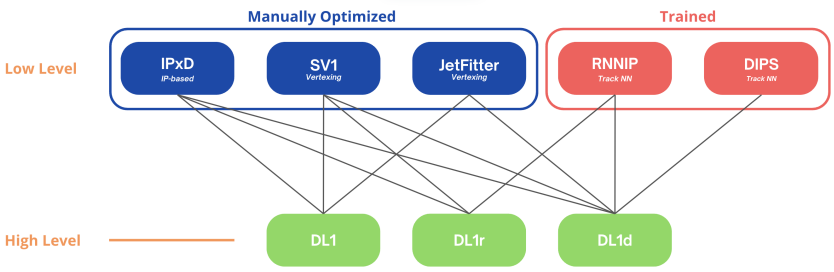


Figure 3.3: Schematic overview of different low level algorithms like IPxD, SV1, JetFitter and the ML based RNNIP and DIPS, whose outputs are fed into the high level taggers DL1, DL1r, DL1d.

In addition to the distinction between low-level and high-level tagging strategies, the identification of heavy-flavour jets requires careful consideration of the kinematic regime in which the decay occurs. The topology of heavy-flavour decays is strongly influenced by the transverse momentum of the parent particle, giving rise to two qualitatively different reconstruction regimes, commonly referred to as resolved and boosted. In the resolved regime, the decay products of an $X \to b\bar{b}/c\bar{c}$ process are sufficiently separated in angle to be reconstructed as two individual small-radius jets, each of which can be independently identified using standard flavour-tagging algorithms. As the transverse momentum of the decaying particle increases, the decay products

become increasingly collimated and are reconstructed as a single large-radius jet, defining the boosted regime.

Tagging algorithms optimized for the resolved topology rely on the presence of well-separated jets and clearly identifiable secondary vertices, and therefore exhibit a marked degradation in performance when applied to large-radius jets. In the boosted regime, the overlap of decay products obscures individual decay signatures and modifies the jet substructure, reducing the effectiveness of small-R jet–based approaches. This motivates the development of dedicated tagging algorithms tailored to large-R jets, designed to exploit substructure information and capture the characteristic features of boosted heavy-flavour decays.

Historically, flavour tagging has been based on a two-stage paradigm, combining low-level reconstruction algorithms with high-level multivariate classifiers. Recent developments in flavour tagging mark a substantial departure from this approach. Modern algorithms move towards end-to-end learning strategies that operate directly on low-level detector information, eliminating the need for manually engineered features.

Within this paradigm, the GN family of algorithms models reconstructed jets as graphs, where nodes correspond to tracks or secondary vertices and edges encode their relational information. These graph-based representations are processed using deep learning architectures capable of learning complex correlations among jet constituents. In particular, transformer-based models, which are neural networks built around attention mechanisms that dynamically weight the relevance of different elements within a set, have emerged as a powerful tool for this task. When applied to graphs, transformers enable the aggregation of information across the entire jet in a permutation-invariant manner, allowing the model to capture both local and global features of heavy-flavour decay topologies.

By directly classifying graph-based jet representations using transformer architectures, these approaches bypass explicit low-level reconstruction and high-level combination steps, resulting in improved discrimination performance and enhanced robustness across both the resolved and the boosted regimes.

Several generations of GN-based algorithms have been developed over the years. As anticipated, the following section introduces the second iteration of this approach, GN2, which is used as a reference point for understanding the underlying methodology and architecture. This provides a natural transition to the boosted variants of the GN family, denoted by the suffix X, which are specifically designed for large-radius jets, about which a detailed discussion of the GN2X algorithm is then presented.

3.2.1 GN2: A Transformer-based Approach to Jet Flavour Tagging

GN2 [31] is a Transformer model that directly processes the low-level inputs of a jet, including the kinematic and topological information of tracks and secondary vertices.

Its approach is depicted in Figure 3.4, where jets are reconstructed using the anti-k_t clustering algorithm with a radius parameter of $R = 0.4$, as implemented in the `FastJet` framework.

Each reconstructed jet is encoded as a fully connected graph, in which nodes correspond to the tracks associated with the jet and are described by a set of features capturing both track-level and jet-level kinematic information. GN2 applies a transformer architecture to this graph representation. Transformers are neural networks based on an attention mechanism that allows the model to dynamically weigh the importance of each node relative to all others. This enables GN2 to learn correlations among tracks and to aggregate information from neighboring nodes in a flexible, data-driven manner, capturing both local and global features of the jet.

The model is trained using Monte Carlo (MC) simulated samples that provide truth labels for jet flavors. To enhance the learning capacity, GN2 employs multi-task learning by introducing auxiliary tasks alongside the main multi-class classification objective. These auxiliary tasks

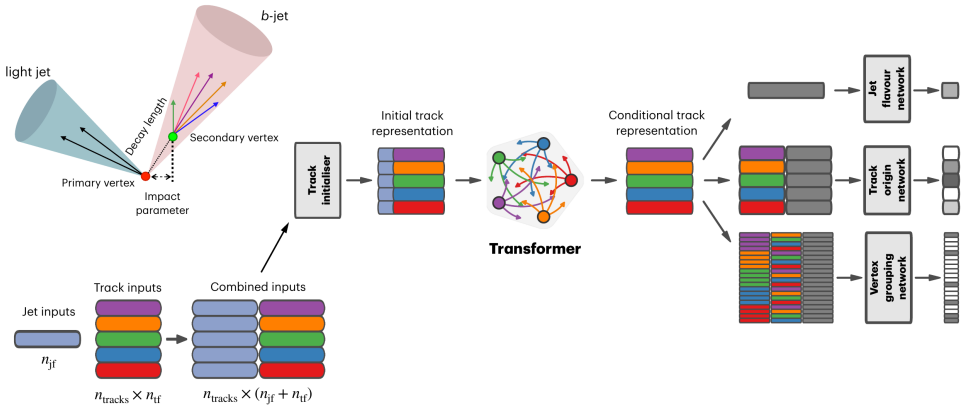


Figure 3.4: Illustration of the GN2 algorithm, showcasing the input variables derived from the jet and its associated tracks, which include properties related to secondary vertices and displaced decay signatures of b-hadrons. The jet features are replicated for each track in the jet to form combined feature vectors, which are then processed by a per-track initialization network. These track representations are subsequently refined through a transformer encoder, capturing the relational and topological information within the jet, resulting in a comprehensive global jet representation. The pooled jet features, along with the output embeddings of individual tracks, are fed into three task-specific neural networks for discriminating between different jet flavors [31].

include vertex classification (e.g., identifying two-track vertices) and track origin prediction, which help the network to better discriminate between primary and secondary interactions and improve the quality of the learned features.

The training optimizes a combined loss function that balances the primary classification loss with the auxiliary task losses, leading to improved generalization and discriminative power. The training dataset comprises simulated jets spanning a wide range of transverse momenta, pseudorapidities, and decay topologies, ensuring that the model can accurately capture the behaviour of both resolved and boosted heavy-flavour jets. This diversity allows GN2 to learn patterns that are representative of realistic collider conditions and improves its robustness across different kinematic regimes.

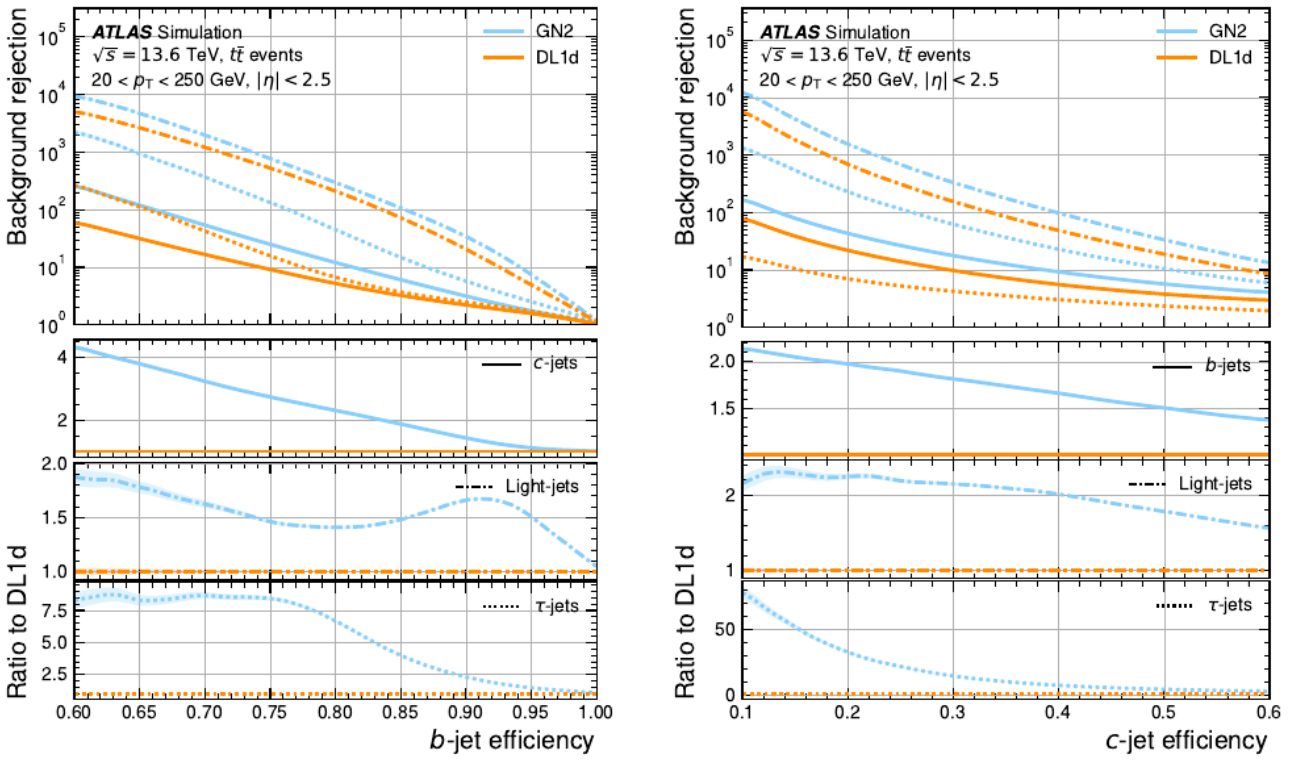
The trained GN2 model produces output probabilities of a jet being a b-jet, (p_b), a c-jet (p_c), a $\tau-$jet (p_τ) or a light jet (p_u). These probabilities are then combined into a single discriminant representing the probability that a jet originates from a b-quark and defined as:

$$D_b = log\left(\frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_u}\right), \tag{3.2}$$

with $f_{c/\tau}$ free parameters determining the relative weighting between the three background scores into the discriminant respectively set to 0.2 and 0.01.

The discriminant value D_b produced by the GN2 modelis used to classify jets as b-tagged or not by applying a set of predefined threshold cuts. These thresholds are determined using di-lepton $t\bar{t}$ events [2]and define different working points, each corresponding to a specific signal efficiency ϵ_b. A jet is then considered as b-tagged, i.e. identified as a b-jet, if its predicted discriminant value exceeds the selected threshold. By adjusting the cut value on D_b, analysts can select working points that balance the b-jet tagging efficiency against the background rejection rate.

[2] b-tagging efficiencies are measured on di-lepton $t\bar{t}$ processes because top-quarks always decays via $t \to Wb$ decay; so requiring both W bosons to decay leptonically yields a highly pure sample of b-jets



(a) (b)

Figure 3.5: GN2 tagger background rejections as a function of the signal tagging efficiency for (a) b- and (b) c-tagging compared to the DL1d ones [31].

After establishing optimized working points, the performance of GN2 is evaluated using simulated Monte Carlo samples, focusing on its capability to reject background jets while preserving a specified signal efficiency. Background rejection, defined as the reciprocal of the mis-tagging rate, i.e. the fraction of background jets incorrectly classified as signal, is used as one of the key metrics. The overall discrimination performance is summarized through the Receiver Operating Characteristic (ROC) curve, which illustrates the trade-off between background rejection and signal efficiency, defined as the number of selected signal jets with respect to the total ones, providing a comprehensive evaluation of the algorithm's effectiveness.

Owing to its fully end-to-end architecture, which does not rely on the outputs of low-level flavour-specific taggers, GN2 learns a generic representation of jet flavour from track- and vertex-level information. As each of the models is trained with three output classes, using it as a c-tagging algorithm is trivially analogous to the approach used for b-tagging with only redefining the discriminant as

$$D_c = log\left(\frac{p_c}{f_b p_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_u}\right), \tag{3.3}$$

with f_b and f_τ respectively 0.3 and 0.01.

The ROC curves, generated using $t\bar{t}$ MC samples, are displayed in Figure 3.5 for both b- and c-tagging efficiencies. In these plot, the GN2 algorithm consistently outperforms the previous DL1d tagger across the entire range of signal efficiency values.

In summary, the GN2 algorithm achieves a significant improvement in flavour discrimination performance with respect to previous taggers, demonstrating the potential of graph-based architectures to effectively exploit the relational and topological information embedded in jet constituents. However, as GN2 is primarily optimized for the resolved-jet regime, its applicability to highly collimated jets is limited, thereby motivating the development of dedicated

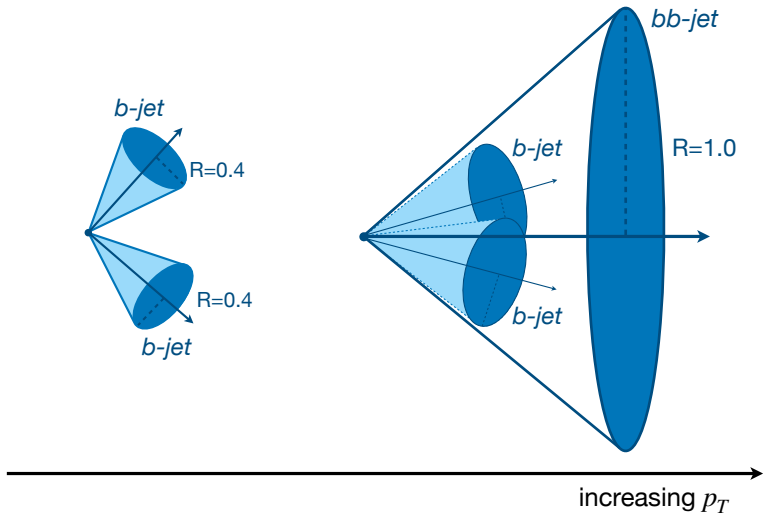


Figure 3.6: Final state reconstruction in resolved and boosted topologies.

tagging strategies specifically designed for boosted topologies.

3.2.2 Boosted Topologies and Dedicated Taggers

The reconstruction of hadronic decays of heavy particles, such as the Higgs boson, top quark, or electroweak bosons, strongly depends on the transverse momentum of the parent particle. As illustrated in Figure 3.6, at low and moderate transverse momenta, in the so-called *resolved* regime, the decay products are emitted with sufficiently large angular separations to be individually reconstructed as distinct, small-radius jets.

However, as the transverse momentum of the parent particle increases, the Lorentz boost causes the decay products to become progressively more collimated. In particular, the angular distance between two decay products, such as the two b-quarks originating from a Higgs boson decay, can be approximated by:

$$\Delta R(b_1, b_2) \sim \frac{m_H}{p_T^H} \frac{1}{\sqrt{z(1-z)}} \sim \frac{2m_H}{p_T^H} \tag{3.4}$$

where m_H and p_T^H are the mass and the transverse momentum of the Higgs boson, respectively, and z denotes the momentum fraction carried by one of the two b-quarks. The expression highlights that the angular separation between the two decay products decreases inversely with the Higgs transverse momentum. For example, an Higgs boson with $p_T^H \sim 200$ GeV will produce two b-jets with angular separation of $\Delta R \simeq 1.25$, whereas for $p_T^H \sim 500$ GeV the separation shrinks to $\Delta R \simeq 0.5$, which is a value comparable to the typical jet radius parameter used in the resolved regime.

As a consequence, in the high-p_T regime, the two b-quarks are often clustered into a single large-radius jet rather than two distinct small-R jets. This *boosted* topology significantly alters the jet's internal structure and challenges conventional flavour-tagging algorithms such as GN2, which are optimized for resolved configurations.

To cope with these challenges, boosted *ah-hoc* tagging techniques have been developed, designed to exploit the substructure of large-radius jets and to identify the presence of multiple, collimated decay products within a single reconstructed object.

The interest in the high-p_T regime is twofold, encompassing both precision studies within the Standard Model and searches for Beyond the Standard Model (BSM) phenomena.

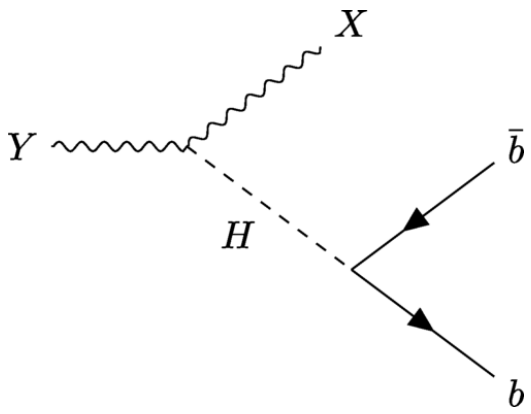


Figure 3.7: Feynman diagram of the target signal process, where the Y is produced in the initial pp collision and decays to a fully hadronic final state via a SM Higgs boson $H \to b\bar{b}$ and a new particle X, assuming that X decay yields a hadronic final state.

From the SM perspective, boosted final states play a central role in several key analyses at the LHC. Processes such as the production of top quark pairs, Higgs bosons, and electroweak gauge bosons frequently populate the high-p_T region, especially at higher centre-of-mass energies.

In this kinematic domain [32], the hadronic decay products of heavy particles become collimated into a single large-radius jet, making boosted reconstruction techniques essential for measurements such as $H \to b\bar{b}$, hadronic $t\bar{t}$ decays, and diboson production.

Accurately resolving the internal substructure of these jets is particularly important for performing differential measurements of Higgs properties, including its transverse momentum spectrum, couplings, and decay branching ratios. Such measurements provide stringent tests of the SM and help constrain higher-order QCD effects, parton distribution functions, and potential deviations due to new physics. Beyond precision studies, they also improve the modeling of SM backgrounds in regions relevant for BSM searches [33].

At the same time, the boosted regime is of particular importance for BSM physics. Many theoretical scenarios predict heavy resonances or new states decaying into SM particles with large transverse momenta.

For example, searches performed by the ATLAS Collaboration target resonances Y decaying into a SM Higgs boson and a new particle X ($Y \to H + X$), as depicted in figure 3.7, focusing on fully hadronic final states where both $H \to b\bar{b}$ and $X \to q\bar{q}$ [34].

In the high-p_T regime, the decay products of H and X become highly collimated and often merge into a single large-radius jet. This produces complex jet substructures, where standard resolved-jet taggers lose efficiency. High-performance boosted tagging algorithms are therefore essential to identify the heavy-flavour content and internal structure of these jets, improving the sensitivity not only for this specific scenario but across a broad class of BSM models that feature boosted heavy particles decaying hadronically.

3.2.3 Xbb: The first ATLAS Boosted Tagger

The first ad-hoc approach to boosted final states classification was the so-called Xbb tagger, introduced by the collaboration in 2020 [35]. The Xbb tagger performs double b-tagging is using both large-R jet and subjets information, and combining kinematic variables associated to the reconstructed large-R jet in the final state with the output of subjets-based high-level taggers, like DL1r, introduced before, or MV2 (better described in Ref.[36]), a Boosted Decision Tree algorithms which combines low-level tagger outputs for classifying small-R jets as b-, c-, or light-jets.

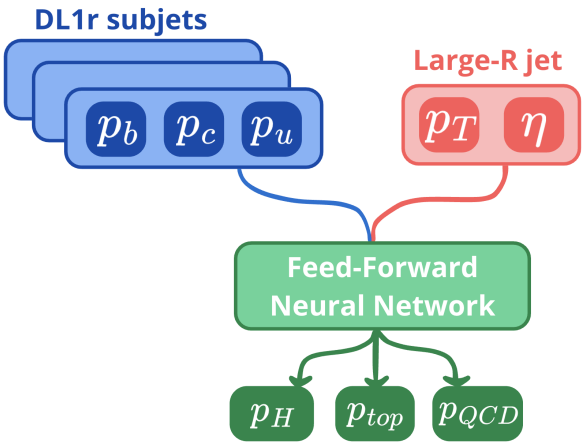


Figure 3.8: Schematic overview of the Xbb tagger. Individual subjets (max. 3) tagging scores obtained with DL1r are used as input together with the large-R jet p_T and η variables. The Xbb tagger then, using as input the DL1r information and the large-R jet ones returns as output the jet probability of being of each category.

In fact, the $H \to b\bar{b}$ decay products are reconstructed as single large-R jets with the anti-κ_t algorithm with R = 1.0, and together with this the b-jets originating from the Higgs candidate decay are reconstructed as VR track-jets with the same parameters introduced in Section 2.3.6. VR track-jets are then ghost-associated [37] to the large-R jets[3].

As depicted in Figure 3.8, the Xbb algorithm consists in a feed-forward neural network in which, using as input the subjets predicted labels by DL1r and the p_T and η values of the reconstructed large-R jet, the returned output is the probability for each event of being an Higgs one (p_H), a top process (p_{top}) or a multijet one (p_{QCD}).

These output probabilities, similarly to what is done in GN2, are then combined into a single discriminant roughly corresponding to a log-likelihood and defined as:

$$D_{Xbb} = ln \frac{p_H}{f_{top} p_{top} + (1 - f_{top}) p_{QCD}} \tag{3.5}$$

with $f_{top} = 0.25$.

The shape of the discriminant is reported in Figure 3.9 for jets in the Higgs, top and multijet samples. Tagging is assigned to a jet whenever either the D_{Xbb} value lies above the selected threshold or the minimum of the two subjet DL1r (MV2) discriminants exceeds that threshold.

The performance in the boosted regime (i.e. asking $p_T > 500$ GeV), are then evaluated by comparing the Xbb tagger with the subjet-based classification obtained using DL1r and two variants of the MV2 tagger. The comparison is carried out in terms of background rejection as a function of the Higgs-jet tagging efficiency.

Figure 3.10 presents the ROC curves obtained for all the tagging configurations, evaluated separately for multijet and top-jet background rejection. Across all cases, the Xbb tagger consistently outperforms the subjet-based approaches, demonstrating higher background rejection at a fixed Higgs-jet efficiency.

Conversely, Figure 3.11 presents the background rejection for both multijet and top-quark processes as a function of the leading large-R jet p_T. The increasing separation between the curves with rising jet transverse momentum illustrates how the performance gap widens in

[3]In ATLAS, ghost association is used to link objects such as tracks or variable-radius track-jets to calorimeter jets by including them in the jet clustering with infinitesimal four-momentum. This procedure does not alter the jet reconstruction, but provides a unique and infrared-safe association between the objects and the reconstructed (e.g. large-R) jets

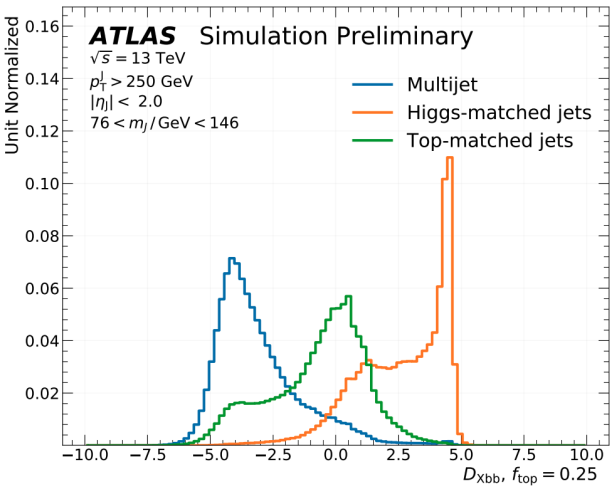


Figure 3.9: Xbb tagger discriminant distribution for each final state cateory [35].

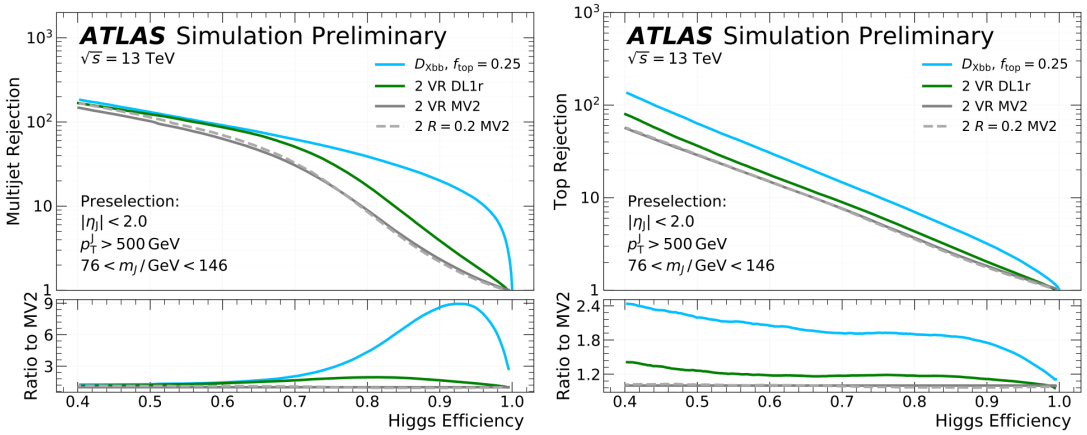


Figure 3.10: Background rejection (multijet on the left, top jet on the right) as a function of the $H \to b\bar{b}$ tagging efficiency, comparing the boosted Xbb tagger curve with respect to the DL1r one and the MV2 one with Variable Radius (VR) jets and Fixed Radius (FR) ones [35].

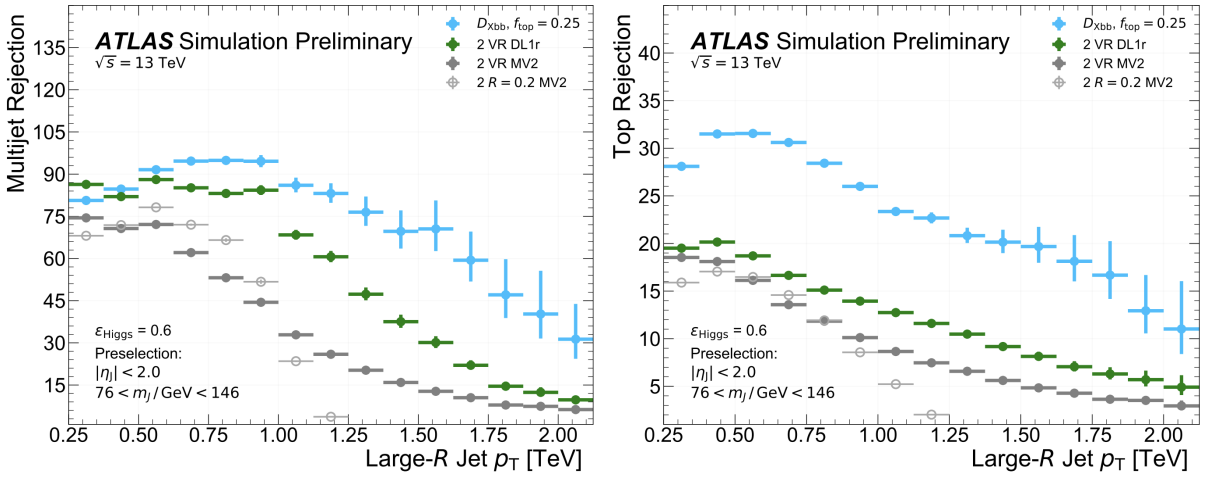


Figure 3.11: Multijet (left) and top (right) jet rejection as a function of large-R jet p_T with a constant 60% Higgs tagging efficiency. Efficiency and rejection are computed with respect to jets that passed the p_T, η and mass preselection requirements [35].

the high-p_T regime. This behaviour indicates that the X_{bb} algorithm is particularly effective for boosted topologies, where the two b-quarks become progressively more collimated and the discriminating power of traditional subjet-based taggers is reduced.

To conclude, the Xbb tagger marked an important milestone as the first boosted $bb-$tagging strategy within ATLAS, achieving strong discrimination power and demonstrating clear improvements with increasing jet transverse momentum. Despite these successes, its architecture still inherits the constraints of traditional taggers, which depend on predefined observables and subjet-based inputs. As jet environments become increasingly complex at higher energies, these limitations become more evident.

This naturally leads to the exploration of modern architectures capable of capturing the full relational structure of jet constituents. The next section introduces GN2X, the state-of-art graph-neural-network–based tagger developed to address these challenges and provide a more flexible and powerful framework for boosted Higgs-jet identification.

3.2.4 GN2X: Employing transformers in the boosted topology

The GN2X $b\bar{b}$ tagger [38] represent the currently used algorithm for identifying $H \to b\bar{b}$ boosted decays within the ATLAS Collaboration. Unlike Xbb, which relies on subjets and pre-defined high-level observables, GN2X operates directly on individual charged-particle tracks, allowing the network to learn correlations and two-prong patterns characteristic of Higgs decays without assuming a specific subjet structure. This fully constituent-level representation is particularly powerful in high-p_T jets, where collimation can obscure secondary vertices and reduce the effectiveness of traditional approaches.

Input Representation and Graph Construction

GN2X takes as input a set variables describing the up to 100 tracks associated to the reconstructed UFO large-R jet with radius R = 1.0. Each track is represented by a feature vector including its kinematic variables, impact parameters and other detector-level properties (listed in Table 3.1) [4].

The tracks are treated as elements of an unordered set. Unlike traditional message-passing graph neural networks that rely on an explicitly defined neighborhood structure, GN2X employs a transformer-style self-attention mechanism. No fixed adjacency is specified: every track can attend to every other track, allowing the network to learn the most relevant interactions within the jet.

By representing the jet as a collection of nodes, GN2X achieves a highly flexible and expressive description of its internal structure. This approach enables the network to naturally identify two-prong patterns typical of Higgs decays, while simultaneously remaining sensitive to more complex configurations that can arise from semi-leptonic decays, additional QCD radiation, or detector effects. In other words, the node-centric representation allows the model to learn a wide variety of substructure features directly from the constituent-level information, rather than being limited to predefined or engineered observables.

[4]The input variables and performance reported here refer to the "standard" GN2X implementation, which is the version used for the calibration studies described in the next chapter and for the boosted $VH, H \to b\bar{b}$ analysis discussed later. Other variants of GN2X have also been explored, including configurations that incorporate subjet-level or flow-constituent information. A detailed description of these alternative implementations can be found in [38].

GN2X Network Architecture and Training

The core of GN2X is built around a series of self-attention layers, inspired by transformer architectures, which operate on the track-level node representations. At each layer, the network computes attention weights between every pair of tracks, effectively allowing each track to "communicate" with all others. This mechanism enables the model to learn which interactions are most relevant for distinguishing signal from background, without relying on any manually defined connectivity.

Following the attention layers, each track representation is processed through feed-forward blocks that refine its latent features. After multiple layers of attention and transformation, a per-jet summary vector is obtained through a permutation-invariant pooling operation, aggregating the information from all tracks into a single representation of the jet. This pooled vector is then passed to fully connected layers that produce the final discriminant outputs, returning the probability for each processed large-R jet to represent an $H \to b\bar{b}$ event, an $H \to c\bar{c}$ one, a top-quark or a generic QCD jet.

GN2X was trained on jets from simulated proton-proton collisions at 13 TeV. Signal jets originate from Higgs boson decays to $b\bar{b}$ and $c\bar{c}$ pairs produced in association with a Z boson (ZH), with training samples generated to produce an approximately flat jet mass distribution. Background jets come from top-quark decays (with hadronic W decays), produced via a hypothetical Z' boson ($Z' \to t\bar{t}$) to achieve a flat jet p_T distribution, and from multijet events generated in slices of particle-level jet p_T to ensure good coverage at high momenta.

Considering the reliance of GN2X classification performance on simulated samples used for its training, an important consideration that emerged during the analysis concerns a campaign-dependent behaviour of the tagger observed in Monte Carlo samples. In particular, one of the MC production periods, corresponding to the 2018 data-taking conditions, exhibited a significantly different tagger response with respect to the others. This discrepancy led to a non-physical data/MC disagreement, as real data were found to be unaffected. The effect was especially pronounced in $t\bar{t}$ samples, where the GN2X discriminant showed a clear and unphysical shift relative to the other MC campaigns.

The origin of this behaviour was traced back to a bug affecting the training samples associated with that specific MC production period, which led the model to learn spurious, non-physical features. This issue was identified at a relatively advanced stage of the analysis, when the optimization of the model working points and the calibration procedure were already close to completion. Due to time constraints and the lack of updated MC productions, a full retraining and correction of the tagger was not feasible within the scope of this work. Consequently, the calibration procedure was finalized using this partially affected configuration, and the corresponding results and methodology are presented in Chapter 4.

While a temporary mitigation, recommended for short-term studies with little top MC usage within the collaboration, consisted in removing the p_{top} component from the GN2X discriminant, but this approach was not suitable for the VH, $H \to b\bar{b}$ analysis presented in Chapter 5, where top-quark processes play a central role both as a dominant background and in the definition of control regions. Instead, an ad-hoc strategy was developed in this work: the training and validation were restricted to MC periods unaffected by the issue, ensuring a stable and physically meaningful tagger behaviour, while for the 2018 dataset, rather than reprocessing new samples, the distributions were rescaled according to the relative luminosity of each year, enabling a reasonable combination across the full Run-2 dataset while maintaining as much internal consistency in the analysis as possible.

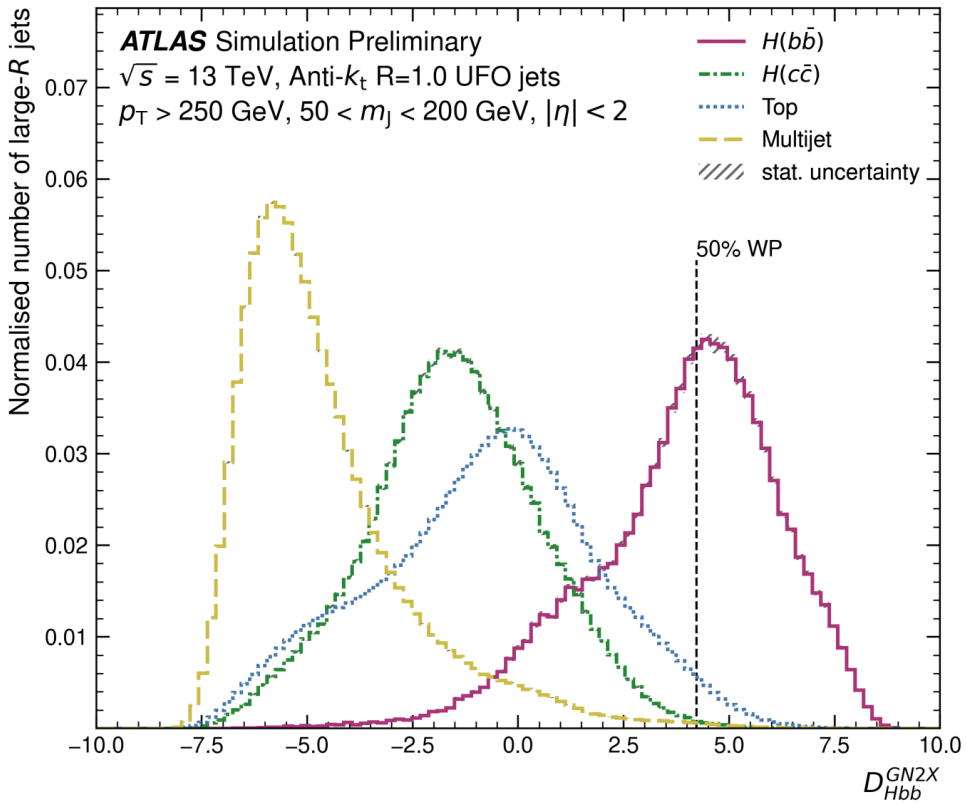


Figure 3.12: GN2X Discriminant shape for each class obtained with $f_{\text{Hcc}} = 0.02$ and $f_{\text{top}} = 0.25$. 50% working point discriminant value is obtained imposing on the $H(b\bar{b})$ jets discriminant distribution to give the corresponding signal efficiency [38].

GN2X Performance

GN2X performance is evaluated, as for the GN2 and Xbb taggers, looking at the rejection values as a function of the signal efficiency. The discriminant for assessing the $H(b\bar{b})$ signal tagging efficiency is defined as a combination of the output likelihoods for each class:

$$D_{\text{Hbb}}^{\text{GN2X}} = ln\left(\frac{p_{\text{Hbb}}}{f_{\text{Hcc}}p_{\text{Hcc}} + f_{\text{top}}p_{\text{top}+(1-f_{\text{Hcc}}-f_{\text{top}})p_{\text{QCD}}}}\right) \tag{3.6}$$

with f_{Hcc} and f_{top} free parameters set respectively to 0.02 and 0.25 for the presented results.

The GN2X discriminant exhibits distinct features for different jet types, reported in Figure 3.12, reflecting the underlying physics and the network's ability to capture constituent-level correlations. For $H \to b\bar{b}$ jets, the output is strongly peaked near high values, indicating that the network confidently identifies signal-like two-prong substructure. Multijet QCD background jets are concentrated at low discriminant values, as expected due to their lack of correlated two-prong features. Starting from the discriminant distribution, several working points are defined corresponding to desired signal efficiencies.

The double-peak structure observed in the top-jet discriminant arises from differences in how the large-R jets capture the top-quark decay products. Jets that fully contain the top decay, including the b-hadron and the hadronic W decay products, tend to exhibit substructure features that partially resemble a Higgs-like two-prong topology, and populate the higher peak of the discriminant. In contrast, large-R jets that do not contain the $b-$hadron from the top decay populates the lower part of the shape.

This distinction highlights the discriminant's sensitivity not just to the overall jet flavor, but also to the internal substructure and constituent-level composition of the jet. It also explains

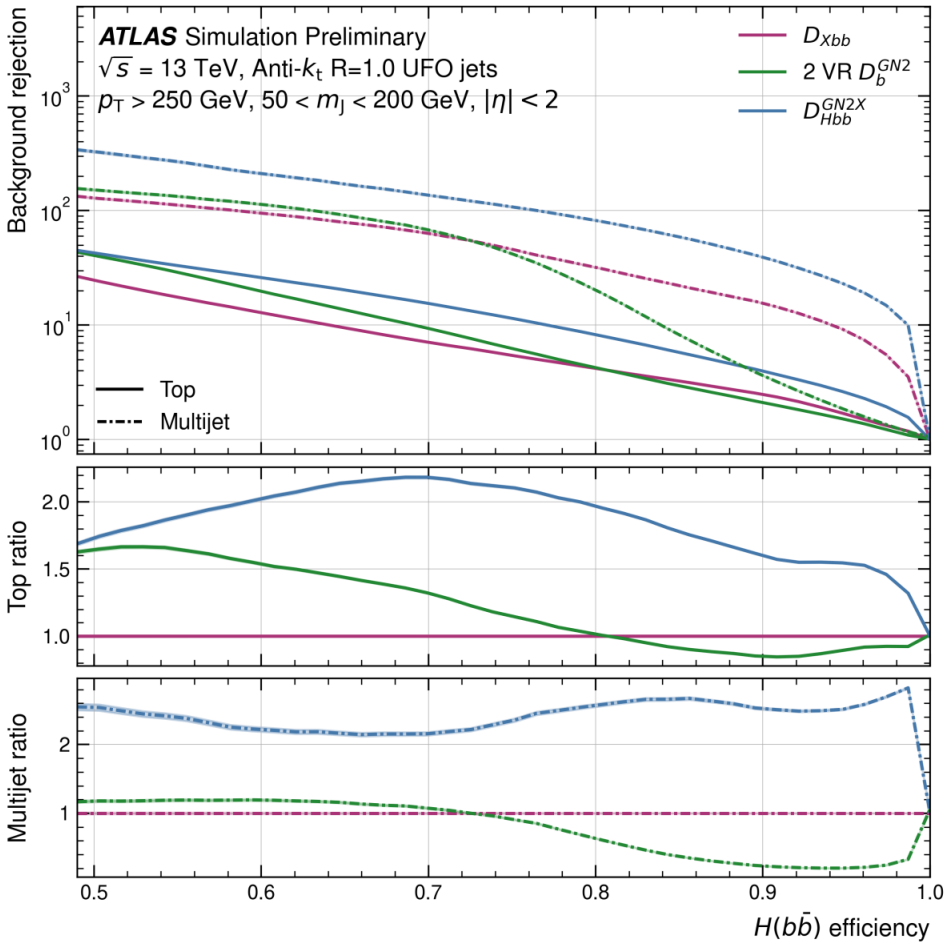


Figure 3.13: Background rejection as a function of the $H(b\bar{b})$ signal efficiency obtained classifying the final state with the GN2X boosted tagger, the previous Xbb one and classifying separately the two reconstructed VR jets with the GN2 tagger optimized for the resolved topology. Sharp variations reflect rejection evaluated only at selected points, especially at high efficiencies [38].

why top-quark rejection is generally lower than multijet rejection at a fixed Higgs efficiency: some top jets can mimic the signal topology closely, while others are easily rejected.

For each signal efficiency, background rejection is computed separately for each background component. The resulting ROC curves are then compared to those obtained using the Xbb tagger on the same evaluation sample, as well as to the state-of-the-art GN2 tagger applied individually to the two VR subjets in a resolved-like topology. Figure 3.13 shows the ROC comparison for $H \to b\bar{b}$ event classification, indicating that, for each background component, the proposed method achieves higher rejection than both the baseline and alternative approaches across the relevant range of signal efficiencies.

Compared to the previous Xbb tagger, the new approach improves rejection at 50% signal efficiency by factors of 1.6 and 2.4 for top and multijet backgrounds, respectively. Performance is also enhanced relative to the 2-tag VR strategy in which the final state is reconstructed with two VR jets classified separately.

As observed in the figure, at low $H(b\bar{b})$ efficiencies, the 2-tag VR subjets baseline slightly outperforms the Xbb tagger, reflecting the benefits of exploiting correlations between the two leading subjets. In contrast, the Xbb tagger uses up to three VR subjets, which improves its efficiency in cases where the b-hadrons are not contained within the leading two subjets.

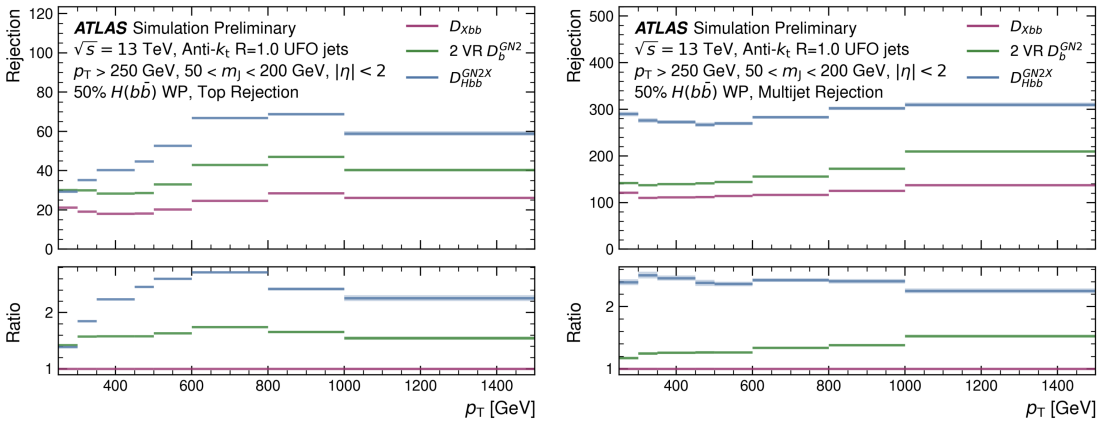


Figure 3.14: Background rejection at 50% of $H(b\bar{b})$ signal efficiency for the top (left) and multijet (right) components [38].

Figure 3.14 shows the background rejection as a function of the large-R jet p_T for a signal efficiency of 50%, separately for the multijet and top-quark contributions. The results are reported for all three models, with GN2X achieving higher rejection than the others across the full p_T spectrum for both background types.

As GN2X outputs likelihoods of the jets being an $H(b\bar{b})$, $H(c\bar{c})$, top or QCD one, with a different definition of the discriminant it can also be used for $H(c\bar{c})$ tagging:

$$D_{\text{Hcc}}^{\text{GN2X}} = ln\left(\frac{p_{\text{Hcc}}}{f_{\text{Hbb}}p_{\text{Hbb}} + f_{\text{top}}p_{\text{top}+(1-f_{\text{Hbb}}-f_{\text{top}})p_{\text{QCD}}}}\right) \tag{3.7}$$

Here, f_{Hbb} and f_{top} are free parameters obtained as for the $H(b\bar{b})$ discriminant, with f_{top}=0.25 as before and f_{Hbb}=0.3.

A comparison is performed only with the GN2 2-VR tagging strategy, as the previous D_{Xbb} tagger was not designed for $H(c\bar{c})$ identification. The corresponding ROC curve, which in this case also includes the non-negligible $H(b\bar{b})$ contamination, is shown in Figure 3.15. The results demonstrate consistently higher background rejection with respect to the baseline across all signal efficiency values, with particularly strong rejection for all background components. These results indicate that the proposed approach could substantially improve the sensitivity of future searches for the yet-unobserved $H \to c\bar{c}$ decay.

The results presented above demonstrate that GN2X achieves state-of-the-art performance in identifying boosted $H \to b\bar{b}$ and $H \to c\bar{c}$ jets, providing higher background rejection than previous approaches such as Xbb or resolved-subjet strategies across the full p_T spectrum. With such high discriminating power, incorporating GN2X into a physics analysis can directly improve sensitivity by enhancing signal selection while suppressing background contributions. In Chapter 5, the impact of GN2X on the boosted $VH(b\bar{b})$ channel will be illustrated, showing the improvement in Higgs signal identification and background rejection enabled by this high-performance tagger.

Meanwhile, in the next chapter, a full description of the calibration methods, which are necessary to make the tagger suitable for physics analyses, will be presented in detail, representing a core part of the work carried out during my PhD.

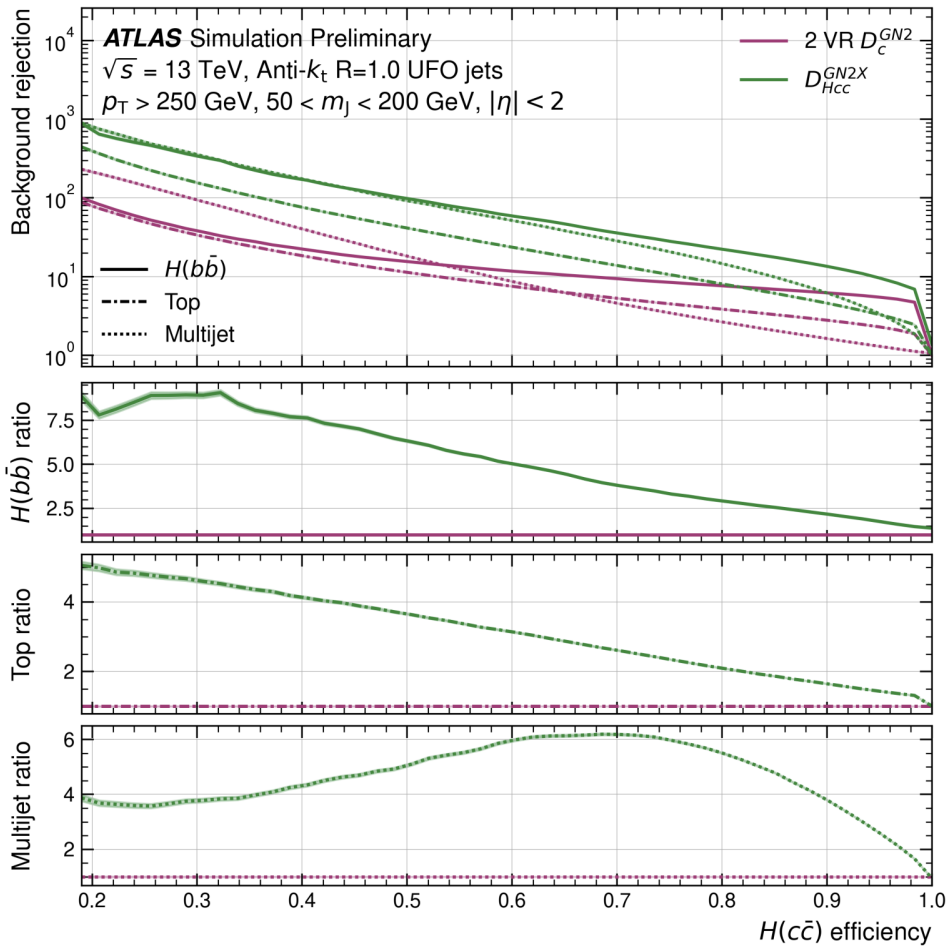


Figure 3.15: $H(b\bar{b})$, top and multijet rejection as a function of the $H(c\bar{c})$ signal efficiency for the GN2X tagger and tagging the two VR subjets with the GN2 tagger (2 VR D_c^{GN2}) [38]

Table 3.1: Input features used by the GN2X tagger. Basic large-R jet kinematics, together with reconstructed track parameters information and constituent hits are used.

Jet Inputs	
p_{T}	Large-R jet transverse momentum
η	Signed large-R jet pseudorapidity
mass	Large-R jet mass
Track Inputs	
q/p	Track charge divided by momentum (curvature measure)
$\mathrm{d}\eta$	Track pseudorapidity relative to the large-R jet η
$\mathrm{d}\phi$	Track azimuth relative to the large-R jet ϕ
d_0	Transverse impact parameter to the primary vertex (PV)
$z_0 \sin\theta$	Longitudinal impact parameter to PV
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on polar angle θ
$\sigma(\phi)$	Uncertainty on azimuthal angle ϕ
$s(d_0)$	Signed transverse IP significance
$s(z_0 \sin\theta)$	Signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits

Chapter 4

Calibrations

To fully exploit the performance of advanced flavour taggers such as GN2X in physics analyses, it is crucial to ensure that their behavior on simulated Monte Carlo (MC) samples matches what is observed in real collision data. Differences can arise from imperfections in detector modelling, limitations in the simulation of physical processes, or reconstruction effects. Without proper corrections, these discrepancies can bias the estimated signal efficiency or background rejection in real collision data, potentially introducing systematic biases in analyses.

During my PhD, I dedicated significant effort to the calibration of the GN2X tagger, with the goal of aligning its performance between MC and collision data, thereby enabling its reliable use in physics analyses targeting boosted $H \to b\bar{b}$ events. The calibration effort was approached chronologically using two complementary methodologies.

In general, the calibration of flavour-tagging algorithms is expressed through scale factors, which are introduced to quantify differences in performance between collision data and simulations. These scale factors determination typically requires data samples with a well-defined and highly pure flavour composition, so that tagging efficiencies can be measured with minimal model dependence. However, the availability of such samples is limited, and for several jet flavour categories, like those dominated by QCD multi-jet production, it is extremely challenging to construct sufficiently pure control regions in data. In these cases, direct data-driven calibrations are either not possible or would carry prohibitive systematic uncertainties, motivating the use of simulation-based approaches to estimate the corresponding efficiencies and uncertainties.

For this reason, the GN2X tagger on multi-jet final states was estimated primarily using Monte Carlo simulations. Nevertheless, information derived from real collision data was incorporated indirectly in the procedure. In particular, data-driven measurements of track-related observables were propagated as systematic uncertainties, allowing potential mismodelling between simulation and data to be accounted for in the calibration.

In this approach, the input variables of the tagger were systematically varied within their known detector and physics uncertainties, and these variations were propagated through the network to evaluate the corresponding impact on signal and background efficiencies. This procedure allows the stability of the mistag efficiency with respect to the modelling of the tagger inputs to be assessed.

The resulting variations in the mistag efficiency were found to be small, leading to scale factors compatible with unity within relatively small uncertainties. This outcome is consistent with a scenario in which the simulation provides a good description of the relevant observables entering the tagger. At the same time, it reflects the fact that the method is intrinsically limited by the set of input variations and systematic uncertainties considered. As a consequence, while this approach provides a useful validation of the modelling and confirms the robustness of the tagger response under known systematic variations, it does not offer a strong handle to probe

potential residual differences between simulation and data beyond those already encoded in the input uncertainties.

For this reason, a complementary data-driven approach was developed for calibrating directly the signal efficiency. This method exploits $Z \to b\bar{b}$ + jets events in data to directly calibrate the signal efficiency of the tagger by comparing GN2X's performance on data and on simulations.

This chapter presents the detailed methodology and results of these calibration efforts, describing both the simulation-based study and the subsequent data-driven calibration. The latter provides the scale factors required for the application of the GN2X tagger in boosted $VH(b\bar{b})$ analyses, which will be described in Chapter 5.

4.1 Adjusted Monte Carlo Method

The Adjusted Monte Carlo method is a calibration strategy designed to evaluate and correct the mistagging performance of flavour-tagging algorithms in situations where direct data calibrations are not feasible or on a longer timescale. It was originally introduced in ATLAS in 2017 to calibrate the mistag rate of light-flavour jets of the MV2 tagger [39], but its underlying philosophy is general and applies naturally to modern taggers such as GN2X.

Flavour-tagging algorithms rely on a wide set of tracking and vertexing observables that encode the distinct properties of heavy-flavour hadron decays, such as displaced tracks, large impact parameters, reconstructed secondary vertices, and the detailed topology of decay chains, as introduced in Section 3.1. Their accurate reconstruction depends critically on the performance of the ATLAS Inner Detector: impact-parameter resolution, alignment, track reconstruction efficiency, and the modelling of rare effects such as fake tracks, interactions with detector material, and long-lived particle decays. Inevitably, some aspects of this detector response are not perfectly reproduced in simulation, leading to mismodelling in the distributions of the tagger inputs and, consequently, in the predicted mistag rate.

The Adjusted-MC method addresses this issue through a bottom-up approach. Instead of correcting the tagger output directly, it examines the individual tracking-related quantities entering the tagger, compares them between data and simulation, and derives variations that bring the simulated distributions into closer agreement with what is observed in data. These variations may include smearing the impact-parameter resolution, modifying the tails of impact parameter (IP) distributions, adjusting fake-track or secondary-track rates, or correcting tracking inefficiencies and detector-geometry effects. Applying these modifications to the reconstructed objects in simulation produces an "adjusted" MC sample that more realistically reflects the detector's tracking performance.

Flavour tagging is then performed on both the nominal and the adjusted MC. The difference in the resulting efficiencies quantifies how sensitive the tagger is to the underlying tracking mismodelling. Combining all relevant variations yields a robust estimate of the correction needed to align the simulation-based mistag rate with the one expected in data, providing the corresponding efficiency scale factors and systematic uncertainties.

Relying only on simulations, this strategy is particularly suited to flavour categories that cannot be calibrated directly in data, such as complex large-R topologies, jets composed of charm-hadron pairs, or mixed heavy- and light-flavour configurations. In these cases, the Adjusted-MC method offers a systematic and physically motivated way to propagate detector-level discrepancies to the final tagger performance.

In practice, the procedure proceeds as follows:

1. Precise measurements of the tracking performance in data and simulation are performed, including impact-parameter resolution and tails, tracking efficiency, fake rate, etc;

2. Implementation of the observed data–MC discrepancies as variations in the MC samples, thereby producing an adjusted version of the simulation;

3. Re-evaluation of the GN2X *bb*-tagger on the adjusted MC to extract the corresponding mistag rate $\epsilon_{ff}^{\text{MC-adj}}$ for each combination of final state flavours ff;

4. Computation of the scale factor for each variation as

$$SF_{ff} = \frac{\epsilon_{ff}^{\text{MC-adj}}}{\epsilon_{ff}^{\text{MC-nominal}}}; \tag{4.1}$$

5. The uncertainties on the input tracking performance corrections are propagated to the mistag rate and subsequently to the derived scale factors, and the envelope of the resulting variations is taken as the associated systematic uncertainty.

Basically the adjusted-MC calibration method relied on data-driven calibrations that are done one step before evaluating the GN2X tagger performance.

The variations used in this work are applied through the `InDetTrackSystematicsAlgs` package, a dedicated Athena-based framework designed to propagate the standard ATLAS tracking systematics to reconstructed tracks. This infrastructure, together with the tools provided in `InDetTrackSystematicsTools`, produces modified track collections in which the relevant systematic variations are explicitly encoded.

Depending on the specific uncertainty under consideration, the tools apply additional smearing to the impact parameters, introduce alignment-related biases, randomly remove tracks to reflect efficiency differences, or adjust the probability of selecting fake tracks. In this way, each systematic variation modifies only the underlying tracking quantities while preserving the remainder of the reconstruction chain unchanged. The resulting adjusted tracks form the basis for the evaluation of the GN2X performance within the Adjusted-MC framework.

Tracking-performance variation name	**Label / short description**
`TRK_RES_D0_MEAS`	d_0 resolution smearing
`TRK_RES_Z0_MEAS`	z_0 resolution smearing
`TRK_RES_D0_MEAS_UP/DOWN`	d_0 smearing up and down
`TRK_RES_Z0_MEAS_UP/DOWN`	z_0 smearing up and down
`TRK_BIAS_D0_WM`	Weak-mode d_0 bias
`TRK_BIAS_QOVERP_SAGITTA_WM`	Weak-mode sagitta bias in q/p
`TRK_RES_D0_DEAD`	Dead modules related bias in d0
`TRK_RES_Z0_DEAD`	Dead modules related bias in z0
`TRK_FAKE_RATE_LOOSE`	Fake-rate with Loose criteria
`TRK_EFF_LOOSE_GLOBAL`	Tracking efficiency (Loose, global material)
`TRK_EFF_LOOSE_IBL`	Tracking efficiency (Loose, IBL material)
`TRK_EFF_LOOSE_TIDE`	Tracking efficiency in dense environment

Table 4.1: Summary of tracking-performance systematic-variation names and their short descriptions.

The list of all the included tracking variations is listed in Table 4.1, and in the next subsection a brief description on how they are computed and why they are relevant for flavour-tagging purposes is given.

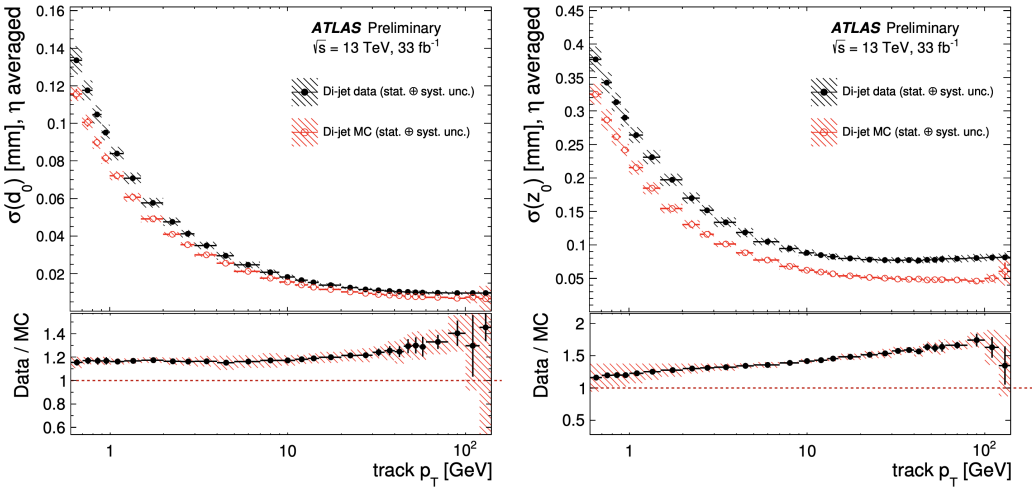


Figure 4.1: Measured track IP resolution in the transverse (left) and longitudinal (right) plane as a function of the track p_T in di-jet events. Systematic and statistical uncertainties are shown in the dashed area and correlations between them are taken into account in the ratio panel [39].

Impact-parameter (IP) resolution systematics The dominant tracking systematic for light-flavour mistagging arises from the modelling of the transverse and longitudinal impact-parameter (IP) resolution. The experimental resolution on the IP of charged tracks, which changes as a function of the tracks p_T and vary from $O(100)\mu m$ to $O(10)\mu m$, is the result of many effects such as the intrinsic resolution of each single hit, the alignment of the tracking components into the detector, the multiple scattering which can happen with the detector material and the accuracy of the track reconstruction algorithms employed. Because these effects are difficult to simulate with perfect accuracy, additional corrections are required to ensure that the IP resolution in MC correctly reproduces that observed in data.

In particular, after a track-to-jet association is performed, tracks are selected according to two identification criteria: the *tight* and the *loose* one. These two selections differ in the minimum number of required silicon pixel and strips hits in the inner detector which are associated to the track, with tight being a sub-set of the loose selection. Also, in the tight criteria, the presence of one hit in the innermost layer of the pixel detector is required, for a final selection that in both cases is meant to represent the various track quality requirements used by the lower-level tagging algorithms.

The track IP resolution is then measured after these selections both on data and MC samples as a function of the track p_T both in its longitudinal (d_0) and transverse (z_0) component.

The corresponding correction applied to MC is the bin-by-bin quadrature difference between the measured resolutions in data and MC:

$$\sigma_{\text{cor}}(d_0) = \sqrt{\sigma^2_{\text{data}}(d_0) - \sigma^2_{\text{MC}}(d_0)}. \tag{4.2}$$

.

Figure 4.1 depict the difference between data and MC samples of the two quantities as a function of the p_T and averaged in the full tracks η range, and the correspondent σ correction factors correspond to the `TRK_RES_D0_MEAS` and `TRK_RES_Z0_MEAS` variations in the Table. The `UP` and `DOWN` variations in the Table are obtained by varying the resolution correction factors by $\pm 1\sigma$, resulting in increased or decreased impact-parameter smearing in simulation.

Weak-mode alignment biases Residual mis-alignments of the inner detector can introduce additional systematic effects on track parameters. In particular, weak-mode deformations, i.e. global alignment distortions that leave the track fit χ^2 nearly unchanged, can produce small

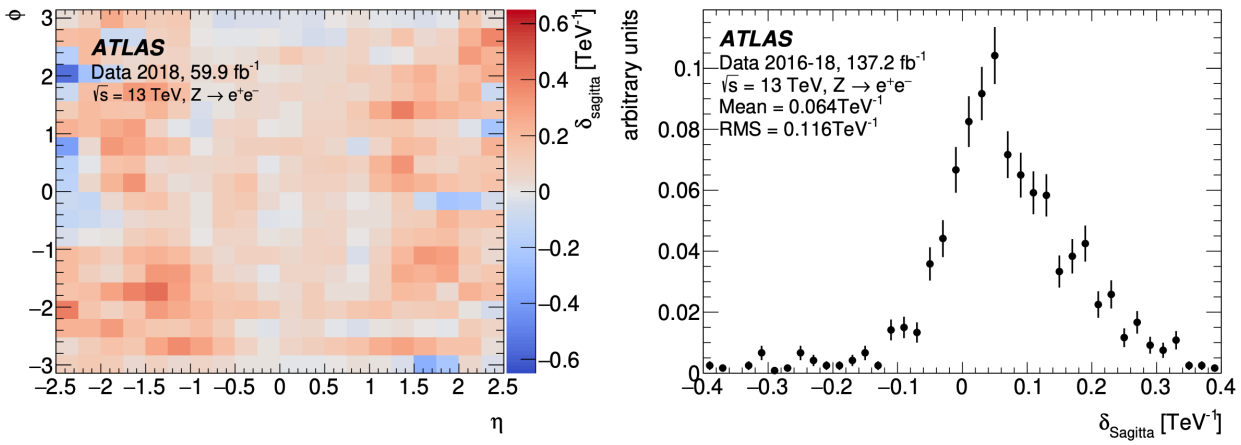


Figure 4.2: Extracted sagitta (curvature) biases as a function of track pseudorapidity η and ϕ (left), obtained from resonance-based momentum-scale studies for Run2 and overall (right) sagitta biases in the Run2 data using $Z \to \mu^+\mu^-$ events [40].

but coherent biases in the transverse and longitudinal impact parameters (d_0 and z_0) or in the track curvature (q/p). These deformations arise from imperfect knowledge of the positions and orientations of the tracking detector elements and can vary across the detector geometry, often depending on azimuthal angle (ϕ), pseudorapidity (η), or radial position.

These effects are invisible to the alignment χ^2, because weak modes preserve the local helical shape of fitted tracks, but they are not invisible to physics observables that depend on the absolute calibration of the reconstructed parameters. In ATLAS, residual weak-mode biases are therefore constrained using dedicated data-driven measurements in addition to the standard track-based alignment procedure.

The systematic bias on the transverse impact parameter is quantified by studying the mean of the reconstructed d_0 distribution for prompt tracks as a function of the track direction. In an ideally aligned detector, the mean transverse impact parameter $\langle d_0 \rangle$ is expected to be consistent with zero across all regions of ϕ and η, within statistical and resolution effects. A coherent shift or a modulation of $\langle d_0 \rangle$ as a function of ϕ or η instead signals the presence of a residual weak-mode deformation.

A complementary and particularly powerful constraint on curvature biases is obtained from two-body decays of resonances with precisely known masses. A distortion affecting the track curvature modifies the reconstructed momentum and therefore shifts the position of invariant-mass peaks. By studying the reconstructed mass as a function of η, ϕ, or track charge, it is possible to identify charge-asymmetric or geometrically dependent shifts that are characteristic of sagitta-like weak modes. Since the true mass of the resonance is known with high precision, the observed displacement can be translated into a residual bias on q/p, providing a direct measurement of the momentum-scale distortion. The corresponding extracted sagitta bias as a function of η is shown in Fig. 4.2, where the residual curvature distortion is quantified and subsequently corrected using $Z \to \mu^+\mu^-$ processes.

It should be emphasized that impact-parameter mean studies and resonance-based momentum-scale measurements represent two of the primary and most robust probes of weak modes in ATLAS, but they are not the only ones. They are used in conjunction with the so called global χ^2 alignment method and additional constraints to ensure that residual biases on d_0, z_0, and q/p are controlled at a level compatible with precision physics measurements [40].

In the adjusted-MC method these variations are encoded as `TRK_BIAS_D0_WM` and `TRK_BIAS_QOVERP_SAGITTA_WM` for weak-mode biases.

Dead modules The presence of dead or inactive detector modules reduces the number of hits available for track reconstruction, degrading the track parameter resolution and potentially shifting the reconstructed IP distributions. Tracks that traverse non-functional regions of the detector have fewer associated measurements, leading to larger uncertainties on the track fit, broader impact parameter resolution, and the possibility of biased residual distributions. Both effects are particularly relevant for flavour tagging, because tagging algorithms rely on precise impact parameter measurements to separate tracks from heavy-flavour decays (displaced) from those originating at the primary vertex (prompt). Even small biases or resolution changes can alter the tails of the IP significance distributions, thereby modifying the mistag probability for light-flavour jets.

In ATLAS, systematic uncertainties related to dead or inactive modules are accounted for by comparing the track reconstruction performance in simulation with and without appropriately modelled inactive modules. The distribution of disabled (dead) modules in simulation is adjusted to match the actual conditions in data, such that the same fraction and spatial pattern of inactive elements are present in both. Differences in track reconstruction efficiency, momentum resolution, and impact parameter performance between nominal and dead-module scenarios are then evaluated; the observed variations are propagated as systematic uncertainties on observables that depend on tracking performance. For example, the degradation in track reconstruction efficiency or resolution due to up to several percent of modules being inactive can be quantified in terms of shifts in the mean and RMS of the impact parameter distributions, and the resulting effects on flavour tagging performance are assessed by re-evaluating the tagging discriminants under these distorted conditions.

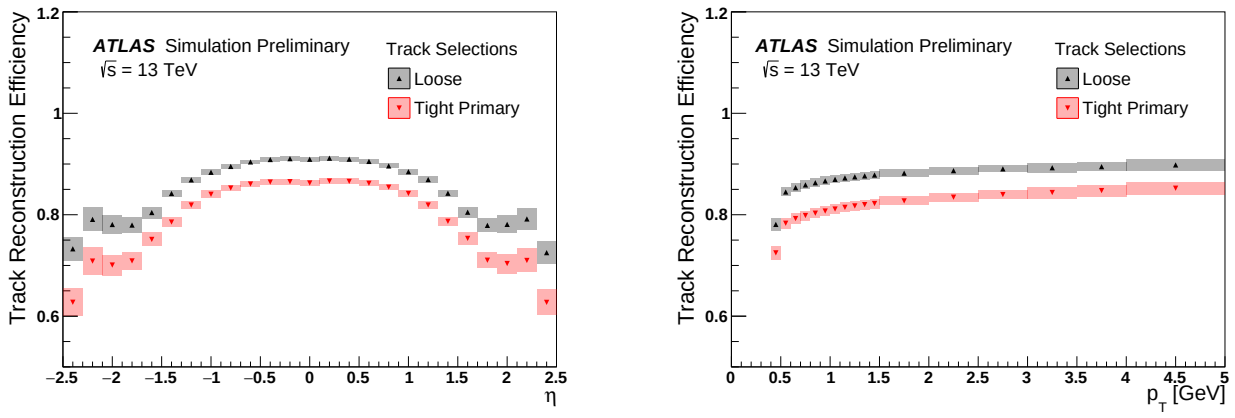
This data-driven simulation tuning ensures that the simulation accurately reflects the inactive detector conditions, and that the resulting systematic uncertainty covers the residual mismodelling between data and Monte Carlo. These systematics are then included in analyses that rely on precise track modelling, particularly those involving displaced tracks such as flavour tagging or secondary vertex reconstruction.

In the adjusted-MC method the IP smearing variations used to account for dead modules is encoded as `TRK_RES_D0_DEAD` and `TRK_RES_Z0_DEAD`.

Tracking efficiency systematics Differences in track reconstruction efficiency between data and simulation are an additional source of uncertainty. These differences are parametrized as maps in (p_T, η) and are derived from dedicated simulation samples where specific detector or physics effects are varied, for example: increasing passive material by 5%, adding 30% more IBL hits, or using alternative Geant4 physics models. The resulting efficiency-loss maps provide a per-track probability for removing a reconstructed track in MC to emulate the observed inefficiency in data, and are presented in Figure 4.3 as reference [41]. These variations are encoded in the `TRK_EFF_LOOSE_GLOBAL/IBL` entries of the Table.

For flavour tagging, reduced tracking efficiency changes the number and distribution of tracks in jets, potentially distorting the impact-parameter patterns used by the taggers or creating artificial secondary vertexes and modelling their fake associated features. Although the effect on light-flavour mistagging is generally small, it is included in the systematic uncertainty to cover possible mismodelling of track-based observables.

Fake-rate systematics Reconstructed tracks that do not correspond to real charged particles —so-called fake tracks— represent an additional source of uncertainty. Fake tracks typically arise from random combinations of hits, pattern-recognition errors, or interactions with detector material. They can distort the impact-parameter distributions used by flavour-tagging algorithms, particularly in the tails, and thus influence the mistag probability of light-flavour jets.



(a) (b)

Figure 4.3: Track reconstruction efficiency as a function of the truth η (a) and p_T (b) values for Loose and Tight Primary track selection, together iwth the total systematic uncertainty associated to each of them. While both uncertainties are displayed in the plot for completeness, only the Loose Primary systematic uncertainty is used in the subsequent studies. The Tight Primary selection is shown for reference and corresponds to a more restrictive track quality criterion. [41]

The rate of fake tracks is measured in both data and simulation, often as a function of the μ value, assuming that the number of "genuine" tracks (i.e. the non fake ones) is to first order proportional to the number of pile-up interactions. Any deviation from linearity is therefore assumed to be due to fake tracks, and the relative deviation from the linear fit is reported in Figure 4.4 for reference [41].Differences between data and MC are then accounted for by dedicated systematic variations, encoded in the `TRK_FAKE_RATE_LOOSE` variation, is measured.

Tracking In Dense Environment and with large-d_0 Tracking performance can degrade in dense environments, such as the cores of high-p_T jets, where the hit density is high and pattern recognition becomes more challenging. The Tracking In Dense Enviroment, shortly as TIDE, variations (`TRK_EFF_LOOSE_TIDE`, `TRK_EFF_RATE_LOOSE_TIDE`) account for the increased probability of losing tracks or creating fakes in these regions, due to hit merging and reconstruction ambiguities. Similarly, large-d_0 efficiency variations (`TRK_EFF_LOOSE_GLOBAL`) address differences in the reconstruction performance of displaced tracks, such as those from secondary decays, which can bias the impact-parameter distributions used by flavour taggers.

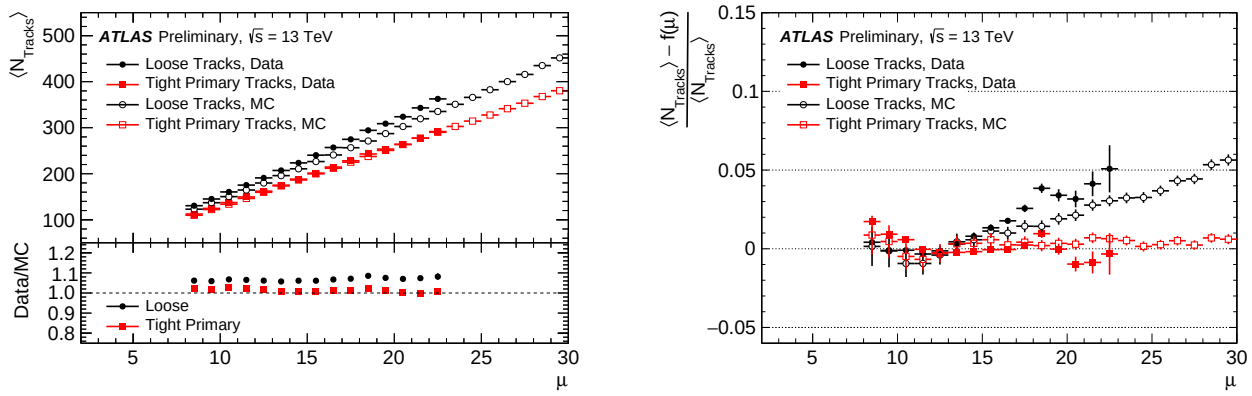
Scale-factor extraction and combination

Each of the tracking-performance variations described above produces an adjusted-MC sample on which the GN2X efficiency is re-evaluated. For each possible flavour composition of the final state ff and every variation i, a corresponding scale factor is extracted as

$$SF_{ff,i} = \frac{\varepsilon_{ff,i}^{\text{MC–adj}}}{\varepsilon_{ff}^{\text{MC–nom}}}, \tag{4.3}$$

with $\varepsilon^{\text{MC–nom}}$ and $\varepsilon^{\text{MC–adj}}$ efficiency measured on nominal and adjusted samples respectively and both associated to an uncertainty $\sigma\left(\varepsilon_{ff,i}^{\text{MC–nom}}\right)$ and $\sigma\left(\varepsilon_{ff,i}^{\text{MC–adj}}\right)$. These individual scale factors constitute the building blocks of the final calibration.

Depending on their origin, in this study the variations enter the combination in two different ways:



(a) (b)

Figure 4.4: Average number of reconstructed tracks as a function of μ for data and simulation applying both Loose and Tight Primary selections computed in the region $10 \leq \mu \leq 15$ (a). These data are then fitted with a linear function $f(\mu) = m\mu$ in the same μ interval, and the relative deviation from the linear fit are used as estimation of the tracking fake rate (b)[41].

- **Data-driven variations.** For variations constrained by data, such as the d_0 and z_0 smearing, whose contributions are computed directly comparing data and MC distributions, the extracted SF_i is treated as a *central value*. To each of these SF_i a statistical uncertainty is added and computed with the standard propagation of the uncertainties

$$\sigma(SF_i) = \sqrt{\left(\frac{\sigma\left(\varepsilon_{ff,i}^{\text{MC–adj}}\right)}{\varepsilon_{ff}^{\text{MC–nom}}}\right)^2 + \left(\frac{\varepsilon_{ff,i}^{\text{MC–adj}}\,\sigma\left(\varepsilon_{ff}^{\text{MC–nom}}\right)}{\varepsilon_{ff}^{\text{MC–nom}\,2}}\right)^2}. \tag{4.4}$$

  As each source of uncertainty is considered independent with respect to the others, data-driven scale factors are then multiplied to form the central value of the final scale factor, while the uncertainties are included into the final combined one.

$$SF_{ff,\text{tot}} = \prod_{i \in \text{data}-\text{driven}} SF_{ff,i} \tag{4.5}$$

- **MC-driven variations.** For variations derived purely from simulation studies (e.g. material-related uncertainties or fake rate ones), the central value is left unchanged. Instead, for each MC-driven variation j, the deviation of the scale factor from unity,

$$\Delta(SF)_{ff,j} = |1 - SF_{ff,j}|\,, \tag{4.6}$$

  is treated as an additional systematic uncertainty and is added in quadrature to the uncertainties of the data-driven components.

Putting all contributions together, the final central scale factor is computed as in Equation 4.5, while its total uncertainty is computed summing under quadrature each contribution.

This prescription ensures that the calibration remains anchored to data whenever possible while incorporating simulation-derived uncertainties where data-driven information is not available.

Because the performance of the GN2X tagger depends strongly on the heavy-flavour composition of the jet, the calibration is performed separately for distinct final-state categories ff. This separation preserves differences in topology, track multiplicity, and displaced-track content that directly influence the performance changes. The categorization of the events is done as a function of the amount of heavy flavour jets in the final state, namely: $0b + 0c$, $1b+ \geq 1c$, $1b + 0c$, $\geq 2b + 0c$, $0b + 1c$ and $0b+ \geq 2c$.

4.1.1 QCD di-jet Monte Carlo samples

The QCD Monte Carlo samples used in this study are produced with the `Pythia 8.235` event generator. As QCD samples cover a large range in cross section, their production is split in several slices in order to efficiently populate the full jet transverse momentum spectrum. Each sample is sliced into 13 regions, according to the leading jet transverse momentum in the lab frame p_T and to the transverse momentum of the outgoing partons in the centre-of-mass frame $\hat{p_T}$. These slices are named `JZ0-12WithSW`, were SW stands for shower weight [42], indicating that each slice have a corresponding weight value defined as the multiplication between the cross section, the filter efficiency[1] and the luminosity, divided by the sum of weights for each slice.

The samples used in this work correspond to the JZ3–JZ5 slices of the inclusive dijet production, whose p_T and $\hat{p_T}$ requirements are reported, together with the associated weight, in Table 4.2.

Slice	Generator	Minimum $\hat{p_T}$ [GeV]	p_T interval [GeV]	Weight
JZ3withSW	`Pythia 8.235`	50	[160,400]	4.5050e+02
JZ4withSW	`Pythia 8.235`	150	[400,800]	1.3797e+01
JZ5withSW	`Pythia 8.235`	350	[800, 1300]	5.3340e-1

Table 4.2: Phytia QCD Monte Carlo slices used in this study, together with the p_T and $\hat{p_T}$ requirements and the corresponding weight. Each weight is defined as cross section × filter efficiency × luminosity [42].

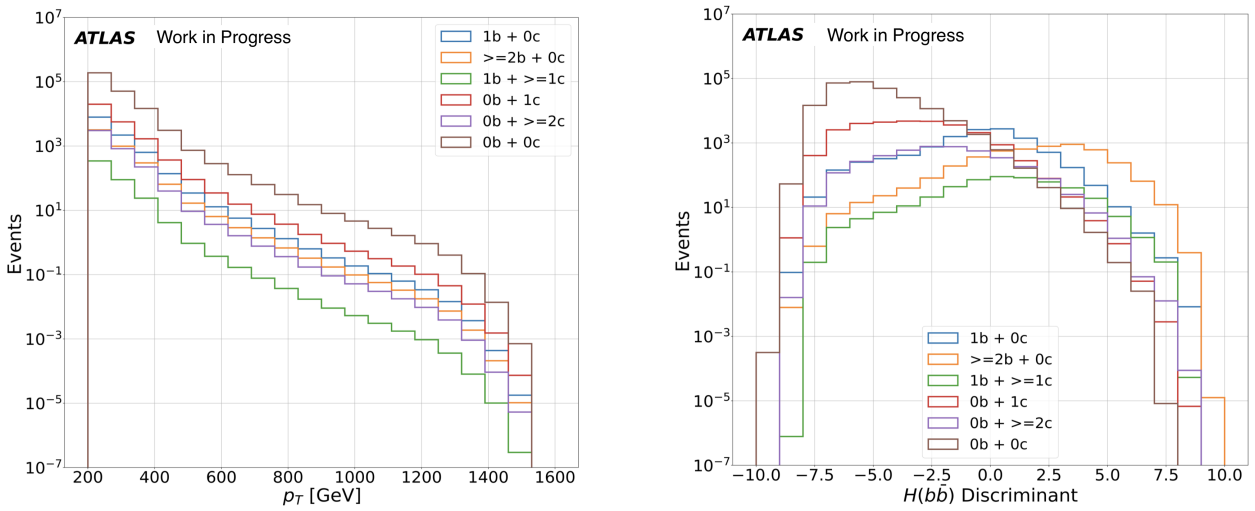
4.1.2 Results

All studies were performed within the `Training Dataset Dumper` (TDD) [43], the ATLAS framework used to produce both the nominal and systematically varied samples used for GN2X training and evaluation. The track-systematic variations provided by the `InDetTrackSystematicsAlgs` package were applied directly inside the TDD, which produced the full set of adjusted tracks. For each event, both nominal and systematically varied, the large-R jets reconstructed and scheduled within the TDD are directly re-tagged with GN2X, yielding as output the corresponding discriminant values for all variations. Jets were associated to tracks through a geometrical matching with $\Delta R = 1.0$, in which the two b-quarks from the Higgs decay are typically reconstructed in the boosted regime.

The GN2X discriminants obtained from the TDD were subsequently used to derive the light-jet mistag efficiencies across a set of working points corresponding to 50%, 60%, 70%, and 85% signal efficiency. For each working point and for every final-state category, the mistag rate was measured as a function of the leading large-R jet transverse momentum. Comparison of each systematically varied sample with the nominal prediction provided the full collection of per-variation scale factors SF_i entering the Adjusted-MC calibration.

Figure 4.5 shows the transverse momentum distribution and the GN2X discriminant distribution, split according to the final-state flavour categories. These results refer to the nominal di-jet Monte Carlo sample, which explains the relative population of the categories: configurations without heavy-flavour jets are naturally dominant, while categories containing charm or beauty jets appear only through mistagging or from secondary heavy-flavour production in the parton shower, and are therefore increasingly suppressed.

[1]Filter efficiency is defined as the ratio between the number of events in each slice passing both the p_T and the $\hat{p_T}$ requirements and the number of events passing only the $\hat{p_T}$ one.



(a) (b)

Figure 4.5: Transverse momentum p_T distribution (a) and GN2X discriminant distribution (b) for the nominal di-jet MC sample, shown separately for the final-state flavour categories used in the analysis. The GN2X discriminant exhibits the expected hierarchy, with heavy-flavour enriched categories shifted toward higher values and light-flavour categories dominating the low-score region.

In the p_T spectra, this hierarchy is clearly visible, with 0b0c events constituting the bulk of the sample and heavier-flavour categories appearing with progressively smaller yields. The GN2X discriminant distributions, as expected, highlight the tagging behaviour of the network. Final states containing true heavy-flavour jets are shifted toward higher discriminant values (as the $H(bb)$ events in Figure 3.12), indicating that the classifier successfully recognizes their characteristic signatures. In contrast, light-flavour and charm-only categories concentrate in the central or lower discriminant regions, consistent with their mistag probabilities.

As a reference for the Adjusted MC procedure, Figure 4.6 compares the GN2X discriminant distributions for the nominal sample and for the adjusted sample in which only d_0-related tracking variations are applied. Although this study is performed on a smaller di-jet sample, the principle is clear: tracking-induced changes in input variables propagate into the discriminant shape. Those shape distortions modify the number of events that pass a given working-point cut and therefore alter the tagger performance.

The distributions in Figure 4.6 are shown separately by final-state category, more precisely with a focus for $ff = bb, cc$, light-light. They demonstrate that d_0 smearing has a substantially larger effect on events without heavy-flavour jets than on heavy-flavour events. Heavy-flavour jets tend to preserve the discriminating features used by the tagger, whereas light-flavour events are more likely to move across the discriminant threshold under the systematic variation. This motivates treating final states separately when deriving scale factors, so that the impact of the Adjusted MC variations is correctly captured for each flavour topology.

Figure 4.7 shows the mistag efficiency as a function of jet p_T for the 0b + 0c light-flavour category, evaluated at the 70% working point of the tagger. The nominal prediction is compared with the results obtained after applying the d_0-resolution smearing variations used in the Adjusted MC procedure. These variations primarily affect the modelling of track-impact-parameter resolutions and are therefore expected to have the largest impact in samples dominated by light-flavour jets, where the discriminant output is most sensitive to small shifts in the input track variables.

As the figure shows, the up and down d_0-resolution variations produce visible changes in

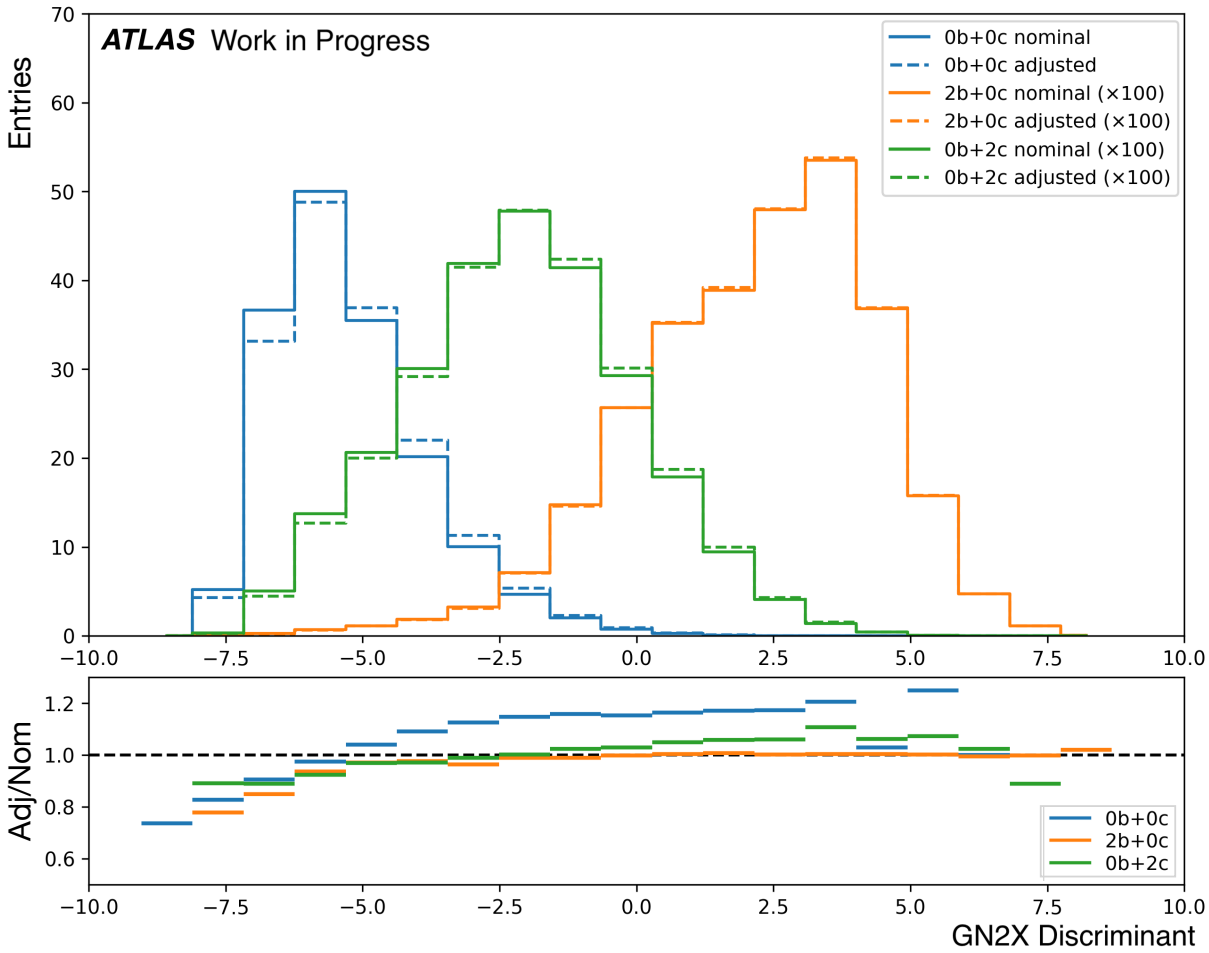


Figure 4.6: GN2X tagger discriminant for nominal samples (solid line) and adjusted one (dashed line) with the $d0$ smearing systematic for different final state categories. Heavy flavour categories are scaled with a factor 100 for a better visualization of the effect of the systematic inclusion into the input samples to the tagger. Lower panel shows the ratio between the shapes before and after the systematic inclusion.

the mistag rate across the full p_T range. The size of the effect increases at high p_T, where the discriminant operates in a more challenging regime and distortions of the d_0 distribution can induce significant migrations across the working-point threshold. This behavior is consistent with the trends observed earlier in the discriminant-shape studies.

It is important to note that the statistical uncertainties of the points are highly correlated across all curves. Each systematic variation is applied to the same underlying MC sample, meaning that the event statistics—and thus the statistical fluctuations—are identical for nominal, up, and down variations. For this reason, when the scale factors are computed, the statistical uncertainty is included only once in the final uncertainty breakdown, while the spread among the systematic variations is treated independently.

Figure 4.8 shows the individual scale factor contributions for the 70% working point, separately for the 0b + 0c and ≥2b + 0c categories. These two categories represent the two extreme cases flavour compositions: the 0b + 0c selection is dominated by light-flavour jets which can be mis-tagged due to input variations, while the 2b + 0c category contains almost exclusively genuine b-jets, whose response is more stable and far from the mistag region.

Because of this difference, the systematics behave very differently. In the 0b + 0c case, variations affecting the modelling of impact-parameter resolutions or the rate of loosely-selected tracks, such as `TRK_RES_D0_MEAS`, `TRK_RES_Z0_MEAS`, and the loose-track-efficiency variations,

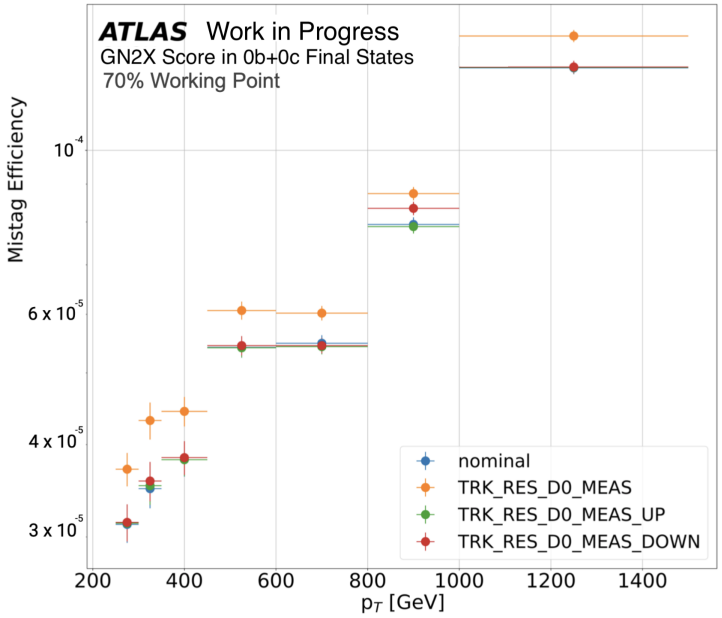


Figure 4.7: Mistag efficiency as a function of jet p_T for the 0b + 0c light-flavour category at the 70% working point of the tagger. The nominal prediction is compared with the d_0-resolution smearing variations used in the Adjusted MC procedure. The statistical uncertainties, shown on each point, are fully correlated between the nominal and varied samples since they originate from the same underlying MC statistics; therefore, the statistical component is included only once in the final scale-factor uncertainty.

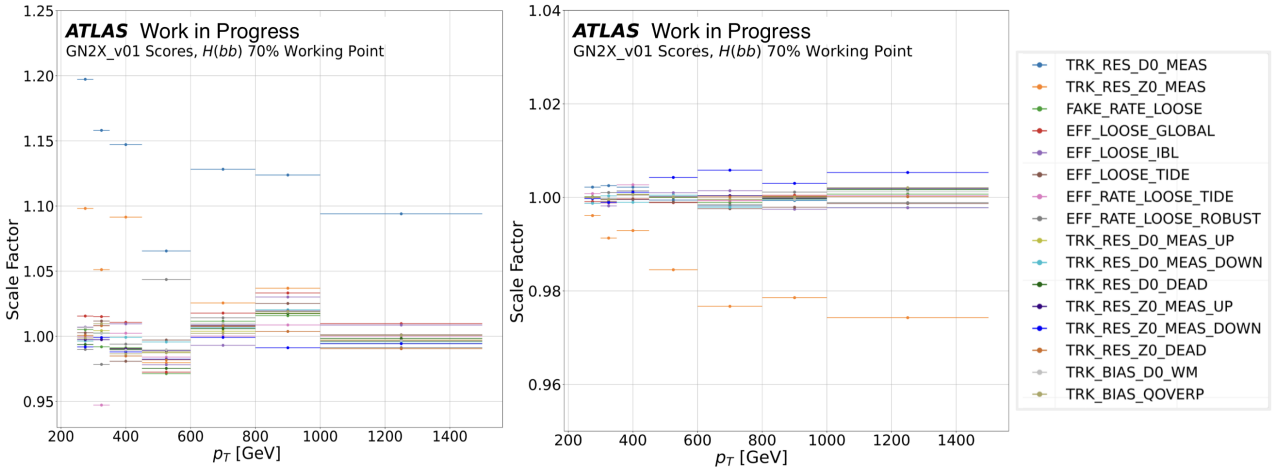


Figure 4.8: Scale factors obtained from the individual systematic variations included in the Adjusted MC procedure, shown separately for the 0b+0c light-flavour category (left) and the 2b+0c heavy-flavour–enriched category (right), at the 70% working point. Each point corresponds to the ratio of the mistag efficiency obtained under a given systematic variation to the nominal prediction. As expected, the 0b+0c category exhibits larger variations due to its strong sensitivity to shifts in the track-impact-parameter modelling, while the $\geq$2b+0c category shows a much more stable response because of the presence of genuine heavy-flavour jets. Statistical uncertainties are not displayed, since all variations are derived from the same underlying MC sample and therefore share fully correlated statistical fluctuations.

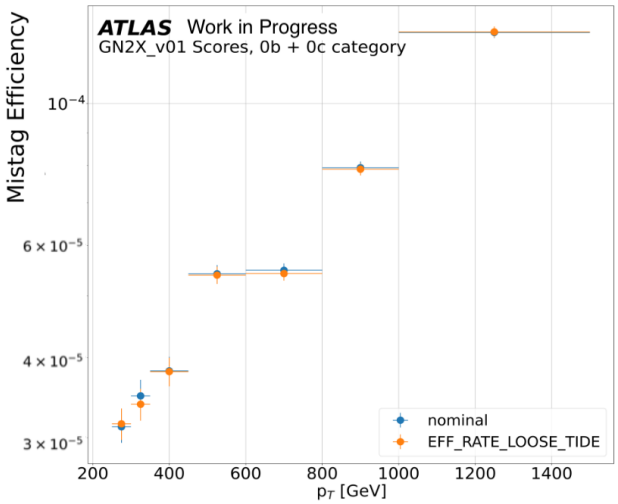


Figure 4.9: Comparison between the nominal mistag efficiency and the `EFF_RATE_LOOSE_TIDE` variation for the 0b+0c category at the 70% working point. A displaced point is observed around $p_T \simeq 300$ GeV; however, deviation is statistically compatible with the nominal prediction. This reflects the limited sample size in this p_T region rather than a genuine modelling effect.

introduce visible shifts in the resulting scale factors across the entire p_T range. The spread seen in the SF panel is therefore a direct reflection of the mistag-efficiency behaviour already observed in Figure 4.7, where the deviations from the nominal efficiency clearly increase with p_T.

In contrast, for the $\geq$2b + 0c category the same variations produce much smaller changes. Here the discriminant output for b-jets sits well above the WP threshold, so modest shifts in resolutions or track-selection efficiencies have a far lower probability of moving jets across the decision boundary. As a result, the scale factors remain extremely stable and close to unity for all systematic variations. The reduced SF spread in this category is therefore a direct consequence of the high intrinsic separation of b-jets from the mistag tail and the limited role of tracking-related uncertainties in this region.

A special case is the `EFF_RATE_LOOSE_TIDE` variation, which shows a slightly displaced point in the 0b + 0c category. A dedicated comparison between the nominal and varied mistag efficiencies, shown in Figure 4.9, confirms that this shift is fully compatible with statistical fluctuations once correlations and the extremely low mistag rates are taken into account. In the mistag comparison plot, adjusted trend follow the nominal one closely, and the apparent deviation in a single p_T bin is well within the expected statistical envelope.

Finally, it is worth stressing that statistical uncertainties are not displayed in the SF breakdown. Each systematic variation is applied to the same underlying MC sample, so the statistical fluctuations of all curves are perfectly correlated. Adding statistical uncertainty independently for each variation would double-count the same effect. Instead, the statistical component is included once in the total SF uncertainty, while the spread of the systematic variations reflects only genuine modelling differences.

The final scale factors obtained after combining the full set of variations are shown in Figure 4.10, where all working points and flavour categories are summarized. This global picture confirms the trends already observed in the per-variation study: only the light-flavour categories, and in particular the 0b+0c configuration, show noticeable migrations with respect to the nominal prediction, whereas categories containing genuine heavy-flavour jets remain stable. This behaviour is consistent with the nature of the available tracking systematics, which predominantly probe the core of the d_0 and z_0 resolutions and therefore have limited impact on the regions most relevant for heavy-flavour tagging.

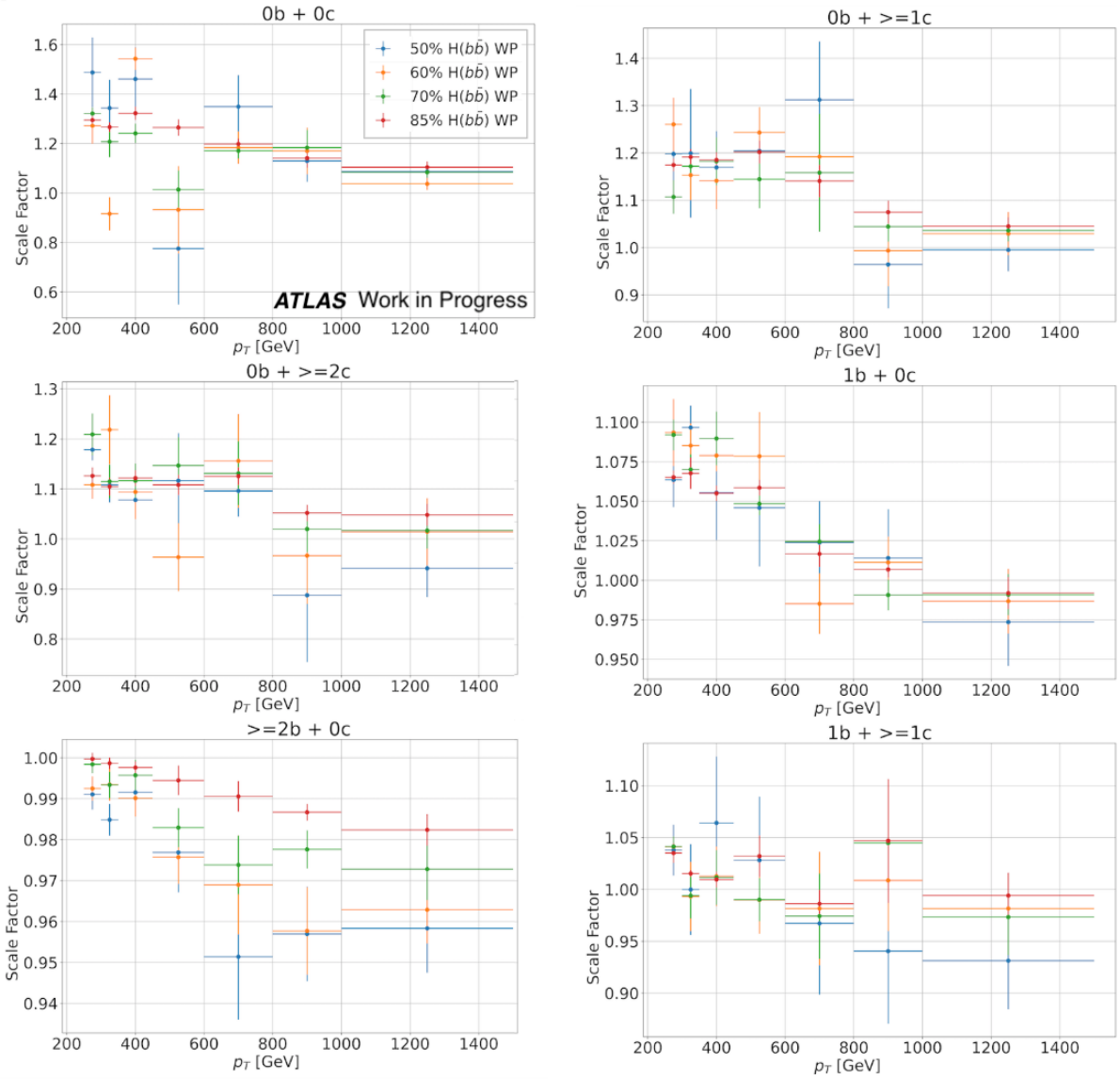


Figure 4.10: Final scale factors for all working points and flavour categories, obtained by combining the contributions of the individual systematic variations according to the prescription described in Equation 4.5. The uncertainties shown include both the statistical component and the combined effect of the systematic variations. Light-flavour categories, particularly 0b+0c, exhibit the most notable deviations from unity, while categories containing heavy-flavour jets remain largely stable due to the reduced impact of track-resolution distortions on genuine b-hadron decay signatures.

In fact, b-tagging algorithms rely heavily on the modelling of the *tails* of the track IP distributions, where contributions from displaced tracks originating from b-hadron decays dominate. Systematic variations affecting these non-Gaussian tails, such as distortions in the long-impact-parameter region, or modifications to the detector material, are not currently included in the ATLAS tracking systematic toolkit used for this study. As a result, the Adjusted MC procedure lacks the variations most capable of inducing sizeable effects in heavy-flavour-enriched samples, limiting its impact in those categories. Further dedicated studies would therefore be required to fully assess its applicability in this regime. Nevertheless, the methodology remains a promising and interesting approach for light-jet calibration contexts.

In the following section, an alternative strategy is introduced to calibrate the signal efficiency of the tagger. Instead of relying on tracking variations applied on MC samples, this approach focuses on the signal in data, making use of a direct data–MC comparison in $Z(b\bar{b})+\text{jet}$ events, where the presence of real heavy-flavour jets enables a robust measurement of the tagging efficiency in data, for calibrating directly the tagger output.

4.2 GN2X Calibration with $Z(b\bar{b})$ + jet events

As discussed in the previous section, the Adjusted Monte Carlo method provides a physically motivated way to estimate light-flavour mistagging uncertainties. However, in its current implementation, it is limited by the absence of systematic variations explicitly targeting the non-Gaussian tails of the track impact-parameter distributions. These tails are precisely the features to which heavy-flavour taggers are most sensitive, and their imperfect modelling in simulation restricts the direct use of the Adjusted-MC approach for calibrating the GN2X signal efficiency.

Possible extensions of the method, incorporating dedicated variations aimed at describing tail distortions or material-related effects, could enhance its applicability to heavy-flavour–enriched samples and deserve further investigation. Such developments, however, lie beyond the scope of this thesis, which instead focuses on a data-driven calibration strategy based on $Z(b\bar{b})$+jets events.

Signal efficiency calibration with $Z(b\bar{b})$+jets events follows the methodology established for previous boosted $X \to b\bar{b}$ taggers ([44], [45]) and extends it to the GN2X architecture. The results presented in this section make use of the full ATLAS Run 2 proton–proton dataset collected between 2015 and 2018 at a center-of-mass energy of $\sqrt{s} = 13$ TeV.

The use of $Z \to b\bar{b}$ events as a proxy for the $H \to b\bar{b}$ signal calibration is motivated by both kinematic and statistical considerations, as Standard Model $H \to b\bar{b}$ signal events cannot be isolated with sufficiently high purity and statistics, particularly at large jet transverse momenta.

Although the Higgs and the Z boson have different masses, the properties that matter for heavy-flavour tagging are closely related. Both particles are colour-neutral resonances that decay into a pair of b-quarks, and in the boosted regime their decay products become highly collimated, forming a single large-radius jet with very similar internal structure, track multiplicity, and secondary-vertex features. Since the GN2X tagger relies primarily on these characteristics, its response to $Z \to b\bar{b}$ events is expected to mirror the behavior it would exhibit for $H \to b\bar{b}$ decays. The strategy therefore assumes that the tagging performance measured on Z bosons is representative of the performance on Higgs bosons. The much larger production rate of Z+jets provides a substantial and clean dataset, which remains statistically powerful even at high transverse momenta. At $\sqrt{s} = 13$ TeV (Run-2) the inclusive Z production cross section around the Z peak is $O(2 \times 10^3$ pb), while the total Higgs production is $O(5.5 \times 10^1$ pb), i.e. the Z yield is roughly 30–40× larger; this abundance allows extraction of finely differential scale factors across the full kinematic region relevant for boosted heavy-flavour tagging.

4.2.1 GN2X Tagger Working Points Derivation

The performance of deep neural network taggers like GN2X strongly depends on how the selection criteria applied to their inputs shape and interact with the underlying event kinematics. A critical challenge that can undermine the validity and precision of such analyses is the phenomenon known as "mass sculpting." Mass sculpting refers to the artificial distortion of the intrinsic invariant mass spectrum of jets, particularly for dominant background processes such as QCD multi-jet production, when a tagging algorithm with non-uniform efficiency across the mass range is applied. This phenomena may cause the mass distribution of the background to resemble that of the signal after the tagger application.

This phenomenon is highly detrimental to physics analyses, as it can lead to false discoveries by creating artificial peaks that mimic signals, compromising the robustness and interpretability of the results.

Traditionally, working points (WPs) for jet taggers were often defined by applying a single, fixed threshold to the tagger's discriminant output. This approach, while straightforward,

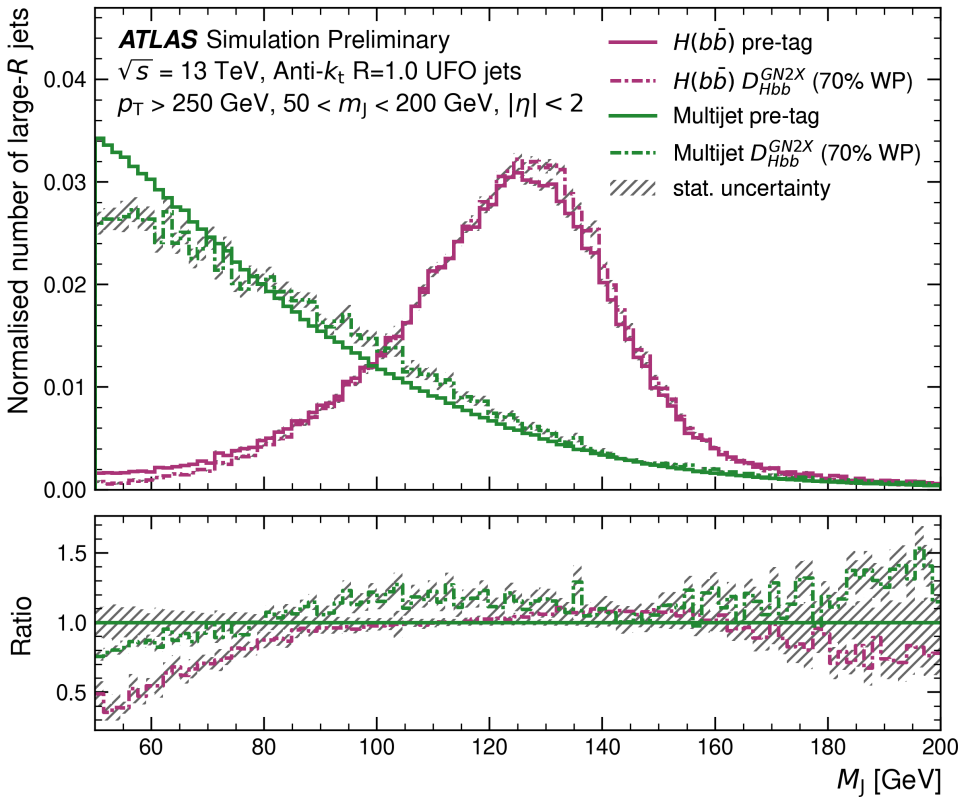


Figure 4.11: Large-R jet mass distributions for $H(bb)$ and multi-jet samples, before and after applying the GN2X tagger cut corresponding to a 70% $H(bb)$ signal efficiency [38].

carries a significant risk of mass sculpting. If the tagger's efficiency varies as a function of the jet's invariant mass (m_J), then a fixed cut will differentially accept or reject jets based on their mass.

This effect is clearly illustrated in the plot of Figure 4.11. This figure displays the normalized jet mass distributions for both the $H(b\bar{b})$ signal (purple) and multi-jet background (green) in Monte Carlo simulation. The solid lines represent the distributions $before$ any tagging, while the dashed lines show the distributions $after$ applying a fixed-cut working point of the GN2X discriminant (D^{GN2X}_{Hbb}) designed for a 70% $H(b\bar{b})$ signal efficiency.

The "Ratio" panel, in which the ratio of tagged to untagged events is reported, shows for multi-jet events a distortion both around the $H(b\bar{b})$ peak and at lower and higher mass values, showing a strong m_J dependence for the tagger performance on the background component. This non-flat ratio confirms that the fixed-cut GN2X working point is not uniformly efficient across the jet mass spectrum for background events, thereby sculpting their intrinsic mass distribution.

To address these challenges, the development of the GN2X tagger has incorporated a new advancement compared to the past: the definition of the so called "Flat-QCD Efficiency" working points. These WPs represent a significant departure from previous fixed-cut approach, which applied a constant threshold to the GN2X discriminant score irrespective of the jet's kinematics.

Instead, the Flat-QCD Efficiency WPs are explicitly designed to be dependent on the jet's kinematics, specifically its invariant mass (m_J) and transverse momentum (p_T). The fundamental principle behind this innovation is to dynamically adjust the discriminant threshold across the phase space such that the efficiency for the dominant QCD multi-jet background remains approximately constant (or "flat") across the relevant jet invariant mass spectrum.

These cut values are established using multi-jet Monte Carlo samples that include recon-

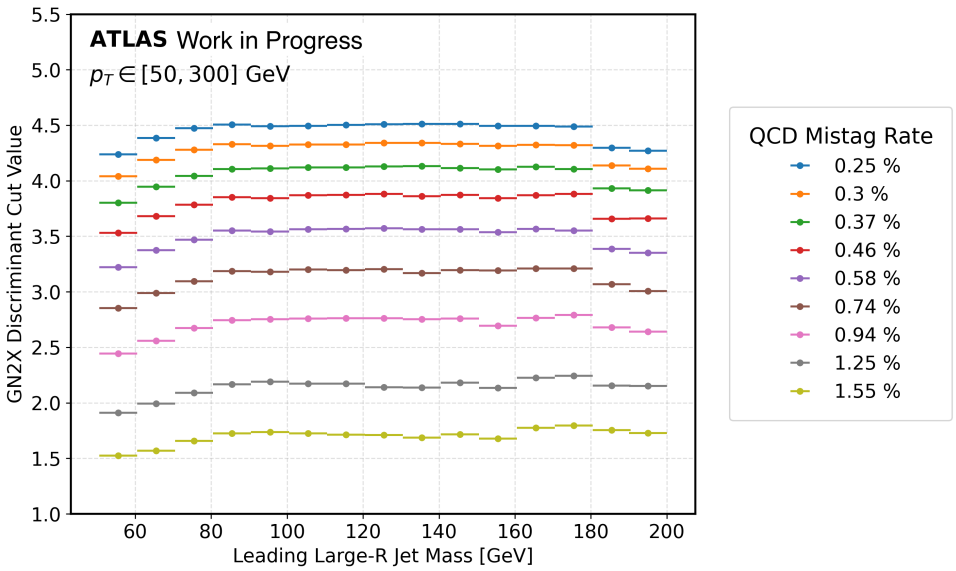


Figure 4.12: Flat-QCD Efficiency GN2X discriminant cut values as a function of the leading large-R jet mass.

structed large-R jets with masses ranging from 50 to 300 GeV, while the p_T is projected in 50 GeV bins over the range of 200 to 1000 GeV. Through this process, the GN2X discriminant cut values are defined as a function of the QCD mistag efficiency, corresponding to the following signal efficiencies calculated on $H(b\bar{b})$ signal samples:

QCD Mistag Rate	$H(b\bar{b})$ Efficiency
0.25%	41.2%
0.30%	45.7%
0.37%	50.3%
0.46%	55.6%
0.58%	60.9%
0.74%	67%
0.94%	72.7%
1.25%	78.7%
1.55%	82.3%

The resulting GN2X discriminant cut values are illustrated in Figure 4.12, clearly showcasing the variable trend of their definition as a function of the leading large-R jet mass. The original (and used in the next sections) discriminant cut values are defined with a granularity of 1 GeV bins in mass, but, for a clearer representation in the figure, the mean values obtained by grouping the mass values into 10 GeV bins are reported.

The success of this kinematic-dependent working point definition is demonstrated in Figure 4.13. This figure directly compares the bb-tagging efficiency as a function of the leading large-R jet mass for both fixed-cut WPs and the new Flat-QCD Efficiency WPs for the GN2X tagger. In stark contrast to the varying efficiencies observed with fixed-cut WPs, the lines representing the Flat-QCD Efficiency WPs exhibit a remarkably stable and flat behavior across a wide range of jet masses.

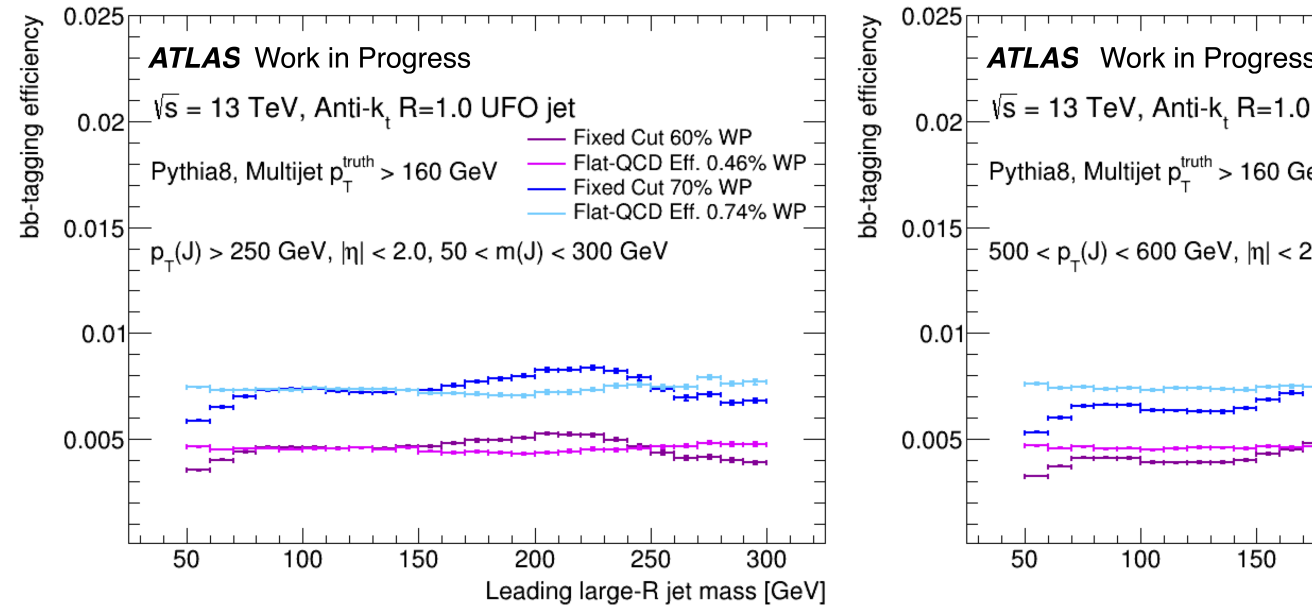


Figure 4.13: Comparison of the GN2X tagger efficiency as a function of mass (left) in inclusive p_T and (right) in $500 < p_T < 600$ GeV region for two working point definitions (with equivalent signal efficiency) in multi-jet simulated events. Fixed cut efficiency (dark purple and dark blue lines) refers to a single cut on the GN2X discriminant for a given efficiency measured in the Standard Model $Z(qq)H(bb)$ simulated sample, for the jets containing the $H \to bb$. Flat-QCD efficiency (magenta and light blue lines) refers to the mass-dependent set of GN2X discriminant cuts to achieve a flat efficiency (in coarse bins of p_T) in the `Pythia8` multi-jet simulated events. Statistical uncertainties of the simulated samples are displayed.

4.2.2 Methodology

The objective of this calibration is to measure the tagging efficiency of the GN2X tagger, defined according to the mass-decorrelated WPs obtained before, for signal jets in both data and simulation. This measurement is crucial for deriving a Scale Factor (SF) that corrects for any discrepancies observed between the efficiencies of the tagger in real collision data versus Monte Carlo (MC) simulations.

The data-to-simulation scale factor is expressed mathematically as:

$$SF(p_T) = \frac{\varepsilon_{\text{data}}(p_T)}{\varepsilon_{\text{MC}}(p_T)}, \tag{4.7}$$

where $\varepsilon_{\text{data}}$ and ε_{MC} represent the efficiencies of the GN2X tagger in data and simulation, respectively. This definition can also be reformulated in terms of event counts, giving insight into the detected and expected signal events:

$$SF(p_T) = \frac{\varepsilon_{\text{data}}(p_T)}{\varepsilon_{\text{MC}}(p_T)} = \frac{\frac{N_{passed}^{\text{sig, data}}}{N_{tot}^{\text{sig, data}}}}{\frac{N_{passed}^{\text{sig, MC}}}{N_{tot}^{\text{sig, MC}}}} = \frac{N_{passed}^{\text{sig,data}}}{N_{passed}^{\text{sig,MC}}} \cdot \frac{N_{tot}^{\text{sig,MC}}}{N_{tot}^{\text{sig, data}}} = \frac{\mu^{post-tag}}{\mu^{pre-tag}}. \tag{4.8}$$

Here, $N_{tot}^{\text{sig, data}}$ and $N_{tot}^{\text{sig, MC}}$ are the total numbers of signal events in data and MC, while $N_{passed}^{\text{sig, data}}$ and $N_{passed}^{\text{sig, MC}}$ are the counts of selected events that pass the GN2X working point criteria.

The μ value, commonly referred to as the signal strength, quantifies the relationship between the observed signal events and the expected signal based on the MC simulation. Mathematically, the signal strength is defined as

$$\mu = \frac{N_{\text{sig}}^{\text{obs}}}{N_{\text{sig}}^{\text{exp}}} = \frac{\sigma_{\text{measured}} \times \text{BR}}{\sigma_{SM} \times \text{BR}_{\text{SM}}}, \tag{4.9}$$

where $N_{\text{sig}}^{\text{obs}}$ is the signal yield extracted from data, $N_{\text{sig}}^{\text{exp}}$ is the expected signal yield from Monte Carlo simulation assuming the Standard Model prediction σ_{SM}, and σ_{measured} is the measured production cross section. In this framework, $\mu = 1$ corresponds to perfect agreement with the Standard Model expectation, while deviations from unity indicate either an excess or a deficit of signal events relative to the prediction.

Within this calibration effort, the post-tag signal strength $\mu^{post-tag}$ reflects the count of tagged events in the collision data as compared to the expected numbers derived from simulations, and in the same way the $\mu^{pre-tag}$ value represents the ratio between data and MC events before the GN2X tagger selection.

The post-tag signal strength is measured with $Z(b\bar{b})$ + jets events, applying on them the GN2X tagger selection and then measuring the differences in signal events fitted on data with respect to the ones that passed the cut on MC simulations.

As directly isolating the $Z \to b\bar{b}$ signal without applying the tagger selection poses significant challenges due to the overwhelming QCD background contribution, the pre-tag signal strength $\mu^{pre-tag}$ is obtained through auxiliary measurements obtained with $Z \to l^+l^-$ events, simply applying the event selection and counting the remaining events.

This approach enables the measurement of the pre-tag signal strength because the leptonic final state selection significantly suppresses the overwhelming QCD background. By focusing on the leptonic decays of the Z boson, the contributions from multi-jet events are effectively mitigated, making it feasible to accurately measure the signal events in both data and Monte Carlo simulations before the GN2X selection.

Events for $\mu^{post-tag}$ measurements are required to have a least one large-R jet identified as Z-boson candidate, and the mass distribution of the leading large-R jet is then used to

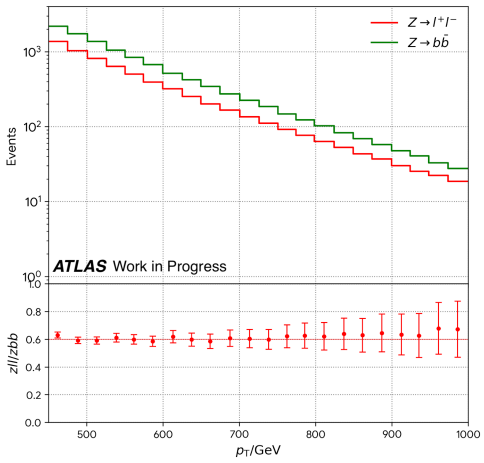


Figure 4.14: Comparison of p_T distributions for $Z \to l^+l^-$ and $Z \to b\bar{b}$ events. The top panel shows the event counts as a function of Z boson p_T, while the bottom panel illustrates the ratio of $Z \to l^+l^-$ to $Z \to b\bar{b}$ events, demonstrating the consistency between the selections applied to each decay channel.

determine the signal strength through a likelihood fit described in the next sections. As this selection results The event selections described in Sections 4.3 and 4.4 for both measurements are intentionally designed to ensure the phase spaces are as similar as possible, employing the same trigger strategy and applying similar cuts to both classes of events.

After having applied the selections, the method's validation relies on the consistency between $Z \to l^+l^-$ and $Z \to b\bar{b}$ events, with their p_T distributions as a function of the Z boson p_T consistent and presented in Figure 4.14 for reference.

Specifically, the pre-tag measurement assumes that the number of observed $Z \to q\bar{q}$ events can be related to the known signal yield of leptonic decays:

$$N^{\text{data}}_{Z\to q\bar{q}} = \mu^{lep} \times N^{\text{MC}}_{Z\to q\bar{q}}. \tag{4.10}$$

Here,

$$\mu^{lep} = \frac{N^{\text{data}}_{ll} - N^{\text{MC}}_{bkg,ll}}{N^{\text{MC}}_{Z\to l^+l^-}} \tag{4.11}$$

quantifies the expected yield of $Z \to l^+l^-$ events in data compared to the MC simulation, adjusted for backgrounds. This relationship, relying on the assumption of having a well-known branching ratio, allows the analysis to bridge the leptonic channel with the hadronic one effectively, assuming that the p_T distributions of both decay modes are similar due to their underlying physics. Thus, the $\mu^{pre-tag}$ measurement relies solely on the leptonic channel.

GN2X Scale Factors are calculated for the 72.7 % signal efficiency (to which we will refer from now on as 70 % for simplicity), in three distinct p_T bins, determined in order to have comparable statistic across all of them: $450 < p_T < 500$ GeV, $500 < p_T < 600$ GeV, and $600 < p_T < 1000$ GeV. Consequently, both the pre-tag and post-tag are measured on events which are divided based on the event leading large-R jet p_T and Scale Factors are obtained independently for each p_T interval.

4.3 $\mu^{pre-tag}$ Measurement

As anticipated, due to the overwhelming QCD background in the non-tagged hadronic channel, a measurement of the $Z(b\bar{b})$ signal strength before applying the tagger is unfeasible. Therefore,

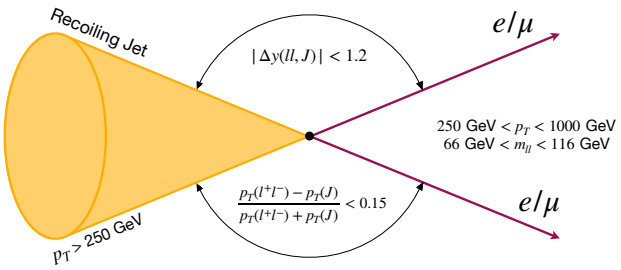


Figure 4.15: $Z(l^+l^-)$+jets event selection criteria.

the pre-tag signal strength is extracted using $Z(\ell^+\ell^-)$+jets events with a simple cut and count analysis in which events are firstly selected in both data and simulations and then the amount of the remaining signal ones is compared in the two cases.

4.3.1 Samples

When selecting $Z(\ell^+\ell^-)$+jets events for the $\mu^{pre-tag}$ measurement, it is essential to understand the various background processes that can mimic the signal ones.

The signal events are $Z \to e^+e^-$+jets and $Z \to \mu^+\mu^-$+jets, while the main background contributions are from diboson processes, including ZZ, ZW and WW events, $W \to l\nu$+jets and $Z \to \tau\tau$+jets. Each nominal component is generated using `Sherpa 2.2.11` [46, 47] generator. To estimate a generator modelling uncertainty, alternative samples are generated with `MadGraph5_aMC@NLO` [48] interfaced with `Pythia8` [49] using the FxFx merging scheme [50]. The following table provides a summary of each background component, together with the nominal and the alternative sample used for quantifying the related modelling systematic uncertainty.

Sample	Description	MC Generator
$Z \to l^+l^-$ + jets	Nominal Signal Sample	`Sherpa 2.2.11`
$Z \to l^+l^-$ + jets	Alternative Signal Sample	`Madgraph + Pythia8 with FxFX`
$W \to l\nu$ + jets	Background	`Sherpa 2.2.11`
$Z \to \tau\tau$ + jets	Background	`Sherpa 2.2.11`
VV (ZZ, WZ, WW)	Background	`Sherpa 2.2.11`

Table 4.3: Summary of MC samples used for the $\mu^{pre-tag}$ measurement.

4.3.2 $\mu^{pre-tag}$ Event Selection

To select $Z(\ell^+\ell^-)$+jets events, the following criteria, schematized in Figure 4.15, are applied:

- Each event must contain at least one large-R jet with $p_T > 200$ GeV.
- The transverse momentum of the dilepton system, $p_{T,\ell\ell}$, is required to lie within the p_T interval defining the analysis bin.
- The rapidity difference between the dilepton system and the leading large-R jet is required to be less than 1.2.

- A symmetry condition on the transverse momentum is imposed:

$$\frac{|p_{T,\ell_1} - p_{T,\ell_2}|}{p_{T,\ell\ell}} < 0.15,$$

  where $p_{T,\ell\ell}$ is the transverse momentum of the dilepton system.

- The reconstructed Z-boson candidate mass must lie within $66 < m_{\ell\ell} < 116$ GeV.

Electron channel Electrons are reconstructed from energy deposits in the tracks matched between the calorimeter and the inner detector, on which several requirements are required both on track properties and quality [51]. Then electrons are identified using a likelihood technique and required to pass a 'medium' working point, additionally requiring isolation by imposing the 'tight' isolation working point [2]. $Z \to e^+e^-$ events are then selected applying additional requirements.

- Each electron must have $p_T > 25$ GeV.

- The transverse momentum balance condition described above is applied to the two electrons.

Muon channel Muon tracks are reconstructed independently in the inner detector and the muon spectrometer systems and required to have a minimum number of hits in each of them and geometrical and momentum matching properties [52]. Then 'medium' identification and 'tight' track-based isolation criteria are applied. $Z \to \mu^+\mu^-$ events are then selected with additional requirements.

- The two muons must have opposite charge [3].

- The same transverse momentum balance and rapidity requirements as in the electron channel are applied.

These selection criteria ensure precise identification of $Z(\ell^+\ell^-)$+jets events, providing a clean sample for the pre-tag signal strength measurement while controlling kinematic distributions and maintaining high purity across the electron and muon channels.

4.3.3 Results: $\mu^{pre-tag}$ values

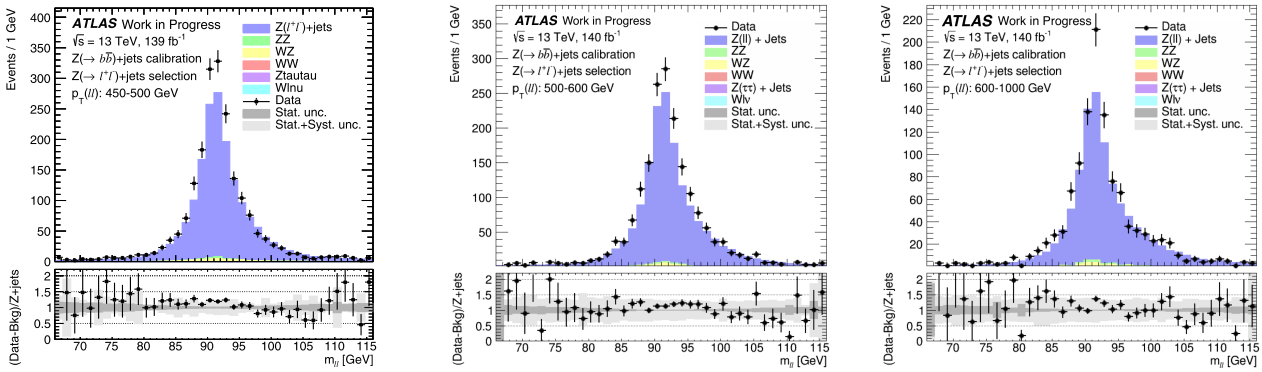
After the events are selected for each p_T range, the measurement of $\mu^{pre-tag}$ is carried out through a straightforward count analysis. In this approach, the number of events in both the data and simulation that fall within the defined mass window is counted, and subsequently, the ratio of these counts is used to determine the signal strength before the GN2X application.

Figure 4.16 report the data and simulation mass distributions after the event selection for each p_T bin. Systematic uncertainties taken into account are, in principle, the ones arising from normalization and factorization scale uncertainties, as well as generator-related uncertainties associated with the modelling of the $Z \to ll$ events.

Measured $\mu^{pre-tag}$ values, together with all the related systematic and statistical uncertainties, are reported in table 4.4.

[2] The 'tight' isolation working point requires the absence of tracks withing the p_T-dependent variable-size cone around the electron.

[3] Unfeasible selection in the electrons final states due to the higher rate of charge mis-identification.



(a) (b) (c)

Figure 4.16: Pre-fit $Z \to l^+l^-$ invariant mass distribution in three Z-boson candidate p_T bins: (a) $450 < p_T < 500$ GeV, (b) $500 < p_T < 600$ GeV, (c) $600 < p_T < 1000$ GeV. All distributions are shown after applying the pre-tag $Z \to l^+l^-$ selection.

p_T [GeV]	[450 - 500]	[500 - 600]	[600 - 1000]
$\mu^{\textbf{pre-tag}}$	1.141	1.139	1.127
Statistical Unc.	± 0.027	± 0.028	± 0.035
Systematics			
Modelling $Z(ll)$+jets	+ 0.1842/-0.1392	+ 0.1882 / -0.1414	+ 0.2495 / -0.1624
Modelling ZZ	± 0.0018	± 0.0020	± 0.0025
Modelling ZW	± 0.0038	± 0.0044	± 0.0056
Modelling WW	± 0.0000	± 0.0000	± 0.0000
Modelling $Z(\tau\tau)$ + jets	± 0.0000	± 0.0000	± 0.0000
Modelling $W(l\nu)$ + jets	± 0.0000	± 0.0000	± 0.0000
Scale	+0.2564 / -0.1769	+0.2532 / -0.1753	+0.2502 / -0.1733
PDF + α_s	+0.0430 / -0.0400	+0.0390 / -0.0365	+0.0414 / -0.0385
Luminosity	+0.0202 / -0.0195	+0.0203 / -0.0196	+0.0202 / -0.0195
Total	+0.3193 / -0.2295	+0.3185 / -0.2291	+0.3563 / -0.2487

Table 4.4: Summary of $\mu^{\text{pre-tag}}$ values along with their statistical and systematic uncertainties.

The measured $\mu^{pre-tag}$ values are found to be stable across the full p_T range in which the tagger is calibrated, with central values around 1.13–1.14, indicating a moderate underestimation of the $Z \to \ell\ell$ + jets normalization in simulation at the level of approximately 13%. No significant p_T-dependent trend is observed, supporting the interpretation of the discrepancy as a global normalization effect. The measurement is dominated by theoretical uncertainties, in particular renormalization and factorization scale variations and Z+jets modeling which will be removed when combining the pre- and post-tag measurements. The stability of the correction across bins ensures that the subsequent extraction of GN2x b-tagging scale factors is not biased by residual pre-tag mismodelling effects.

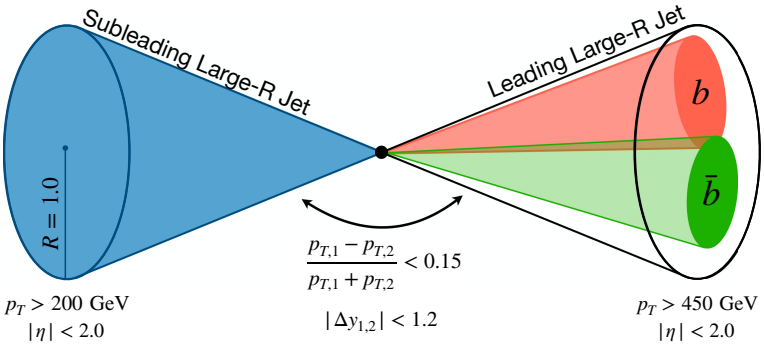


Figure 4.17: $Z(b\bar{b})$ + jets event selection criteria.

4.4 $\mu^{post-tag}$ Measurement

The $\mu^{post-tag}$ value is obtained by fitting the observed leading large-R jet mass data distribution with a signal-plus-background (S+B) template and comparing the resulting signal yield to the corresponding prediction from simulations. This procedure allows for the extraction of the signal strength while accounting for contributions from both signal and background processes.

4.4.1 Samples

When selecting $Z(b\bar{b})$ + jets events the background contributions to be taken into account arise from $W \to qq$ events, where the final states cs or cd jets may be mistagged as $b-$jets, $Z \to qq$ events, which constitute an irreducible background for the $Z(b\bar{b})$ signal, $t\bar{t}$ production, where each top quark decays to a $b-$quark, naturally yielding a $bb-$final state and QCD multi-jet events in which the final state light jets are mis-identified as signal ones.

Sample	Description	MC Generator
$Z \to q\bar{q}$ + jets	Nominal Signal Sample	`Sherpa 2.2.11`
$Z \to l^+l^-$ + jets	Alternative Signal Sample	`Madgraph + Pythia8 with FxFX`
$W \to q\bar{q}$ + jets	Background	`Sherpa 2.2.11`
$t\bar{t}$ + jets	Background	`Sherpa 2.2.11`
$H \to b\bar{b}$	Background	`Powheg + Pythia8`
multi-jet	Background	`Pythia8`

Table 4.5: Summary of MC samples used for $\mu^{post-tag}$ measurement.

Table 4.5 shows that, while QCD events are generated using `Pythia8`, all other samples use the same MC generators adopted for the $\mu^{\text{pre-tag}}$ measurement. The $H \to b\bar{b}$ sample is included only for preliminary studies and is listed in the table because, in principle, it can constitute a background in the selection of $Z(b\bar{b})$ events.

4.4.2 $\mu^{post-tag}$ Event Selection

Hadronic events used in the measurement are selected according to the following criteria, also depicted in Figure 4.17:

Large-R Jet Requirements UFO Large-R jets with radius parameter $R = 1.0$ are used, each reconstructed large-R jet is selected if it passes the following kinematic cuts:

- Leading large-R jet: $p_T > 450$ GeV, $|\eta| < 2.0$.
- Sub-leading large-R jet (if present): $p_T > 200$ GeV, $|\eta| < 2.0$.
- Symmetry and rapidity constraints to ensure a well-balanced dijet topology:

$$\frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}} < 0.15, \quad |\Delta y_{1,2}| < 1.2$$

  where $p_{T,1}$ ($p_{T,2}$) is the transverse momentum of the leading (sub-leading) jet, and $\Delta y_{1,2}$ is their rapidity difference.

Z-boson Candidate Selection The Z boson candidate is reconstructed in the fully hadronic boosted regime as the leading large-R jet in events containing at least two such jets. The identified leading large-R jet is required to have a p_T value in the $[450, 2000]$ GeV, and a reconstructed mass in the $[50, 300]$ GeV range. Z boson candidate events used for the $mu^{post-tag}$ extraction through a likelihood fit are selected applying the GN2X Flat-QCD efficiency cuts corresponding to the 70% of signal efficiency defined in Section 4.2.1. In simulated events, the flavour of each of these track-jets is done by geometrically matching the jet to 'truth' hadrons, using 'truth' information from the generator's event record, allowing them to be classified as originating from b-, c-, or light-flavor quarks.

Alternative working points, corresponding to tighter or looser signal efficiencies, have been studied to evaluate the stability of the measurement and the impact of the tagging requirement on both signal efficiency and background rejection. However, the 70% WP is chosen as the nominal configuration, as it provides an optimal compromise between retaining a sufficiently large $Z(b\bar{b})$ signal yield and achieving strong suppression of the dominant QCD background.

Tighter working points improve background rejection at the cost of reduced signal statistics, while looser working points increase signal acceptance but result in larger background contamination. The 70% WP therefore maximizes the overall sensitivity of the $\mu^{\text{post-tag}}$ extraction and is used throughout the nominal analysis.

4.4.3 S+B Template Definition

The $Z(b\bar{b})$ signal template used in the signal-plus-background fit is obtained directly from the inclusive $Z \to q\bar{q}$ MC sample. Events are reconstructed according to the hadronic selection described before, ensuring that the kinematic properties of the jets match the analysis phase space.

To isolate $Z \to b\bar{b}$ decays, the GN2X tagger is applied to the leading large-R jet, and events passing the 70% working point are taken as signal. This procedure effectively selects the subset of inclusive Z events corresponding to hadronic decays containing two b-quarks, while preserving all correlations and detailed shape features present in the MC.

In addition to the signal template, a detailed description of the background component is required to construct the full signal-plus-background model used in the fit. The background template is built to accurately capture the dominant contributions entering the hadronic selection, most notably multi-jet QCD production and other minor processes. In the following sections, the strategy adopted to characterise the background will be presented step by step: starting from the methodology used to identify the main background sources, then describing how their shapes are extracted or modelled, and finally detailing how the full background template is constructed and incorporated into the fit.

4.5 Background Modelling

The starting point in defining the background template is to quantify the role of the various background processes and establish which ones must be included explicitly. The sensitivity values in Table 4.6 (computed as $S/\sqrt{B}$ with respect to the QCD background) provide a quantitative assessment of how each background component impacts the extraction of the $Z(b\bar{b})$ signal within the GN2X flat-mass 70% working point. Sensitivities are computed in the final S+B mass-window between 50 and 150 GeV, chosen in order to reduce the $t\bar{t}$ peak above. As reported also in Table 4.6, the $Z \to q\bar{q}$ process exhibits the highest sensitivity among the non-QCD backgrounds, with values around 17 across all p_T intervals. This reflects the fact that $Z \to q\bar{q}$ events represent an irreducible background as they contain genuine two-prong substructure and populate the same jet-mass region as the $Z(b\bar{b})$ signal, making them particularly "signal-like" at the event level.

However, despite their large per-event sensitivity, the *absolute yield* of $Z \to q\bar{q}$ events remains much smaller than that of the overwhelmingly dominant QCD multi-jet background. In contrast, the sensitivities for $t\bar{t}$, $W \to q\bar{q}$, and $H \to b\bar{b}$ remain at or below unity, demonstrating that their contribution in the 50–150 GeV mass range is negligible for the purposes of extracting the $Z(b\bar{b})$ signal.

For this reason, the background model in the S+B template concentrates on an accurate description of the QCD component, which dominates both the shape and normalization of the background beneath the $Z(b\bar{b})$ peak. Including $Z \to q\bar{q}$ as a separate background would offer no practical benefit: its yield is too small to alter the mass distribution significantly, and its peak overlaps with the $Z(b\bar{b})$ signal, potentially introducing fit degeneracies.

Table 4.6: Inclusive Sensitivity Values for Different Background Components

Sensitivity ($S/\sqrt{B}$)	[450, 500] GeV	[500, 600] GeV	[600, 1000] GeV
$t\bar{t}$	1.50	1.12	1.21
$W \to q\bar{q}$	0.54	0.54	0.83
$Z \to q\bar{q}$	17.38	16.77	17.55
$H \to b\bar{b}$	0.61	0.61	0.37

To develop a reliable description of the QCD background, a sequence of dedicated studies was carried out to identify the most robust modelling strategy and to minimize potential biases in the final S+B fit. The following sections provide a structured overview of this workflow, which formed a core component of my PhD research.

The procedure begins with a data-driven reweighting of the QCD dijet template using the sidebands, implemented to correct the poor data-MC agreement observed in the leading large-R jet mass distribution. Once an improved baseline description is obtained, the next step is the selection of an appropriate analytical function through a dedicated spurious–signal test. After introducing the conceptual definition of this test, its practical implementation is discussed in detail.

In practice, applying the spurious–signal procedure directly to the reweighted QCD templates is unfeasible, as the reweighting induces significant bin-to-bin fluctuations. To overcome this limitation, the distributions are smoothed using a Gaussian Process Regressor (GPR). A brief explanation of the GPR algorithm and its smoothing properties is provided to clarify its role in producing stable templates suitable for functional testing.

The smoothed QCD spectra must then be made compatible with the fitting framework, which requires unweighted entries rather than continuous weights. For this reason, the GPR output is discretised ("rounded"), and new statistical uncertainties are assigned to each bin according to a Poisson-like prescription.

As a poissonian description of the errors represents an overestimation with respect to the intrinsic behaviour of the smoothed ditributions, a dedicated statistical approach was adopted. Instead of testing the function on a single rounded template, the spurious–signal test is performed on ensembles of pseudo-data. Specifically, 200 toy datasets are generated around the reweighted, smoothed, and rounded QCD distribution. For each toy, the S+B fit is repeated and the extracted spurious–signal parameters (μ and its uncertainty σ) are stored. The resulting distributions of μ and σ provide a robust basis for identifying the functional form that minimises the spurious signal and therefore offers the most reliable description of the QCD background.

4.5.1 Dijets reweight

The QCD reweighting procedure serves to significantly improve the agreement between experimental data and MC simulations for the multi-jet background, a critical step for accurate signal extraction in particle physics analyses. The procedure is data-driven and begins by defining a specific ratio that quantifies the discrepancy between observed and simulated QCD contributions.

The quantity of interest is the ration between Data*, explained better in the next lines, and the MC dijets contribution, calculated per bin across the dijets mass spectrum. Data* is a quantity representative for the number of data events primarily originating from QCD multi-jet production, obtained by subtracting contributions from other well-modeled non-QCD backgrounds from the total observed data. Specifically, Data* is defined as:

$$N^{\text{Data*}}[i] = N^{\text{Data}}[i] - N^{\text{MC}}_{\text{Wqq}}[i] - N^{\text{MC}}_{\text{ttbar}}[i] - N^{\text{MC}}_{\text{Zqq}}[i] \tag{4.12}$$

where, for each bin i in the [50,150] GeV mass fit range, N^{Data} is the total number of observed events observed in data distributions, and $N^{\text{MC}}_{\text{Wqq}}$, $N^{\text{MC}}_{\text{ttbar}}$, and $N^{\text{MC}}_{\text{Zqq}}$ are the simulated event counts for $W \to q\bar{q}$, $t\bar{t}$, and $Z \to q\bar{b}$ processes, respectively. These subtractions remove known non-QCD physics processes, allowing isolation of the data component analogous to the dijets MC sample, which is the simulated QCD multi-jet contribution.

Thus, assuming $\mu = 1$ for each process, the ratio being analyzed for each bin is:

$$\frac{\text{Data*}}{\text{MC}_{\text{dijet}}} = \frac{\text{Data} - \text{MCWqq} - \text{MCttbar} - \text{MCZqq}}{\text{MC}_{\text{dijet}}} \tag{4.13}$$

examined in the sidebands regions [50,70] GeV and [110,150] GeV in order to ensure dominance by the continuous QCD multi-jet background, critically avoiding the Z-boson mass peak.

The Data* to MC_{dijet} ratio distribution is then fitted into the sideband regions with polynomial functions of varying degrees N

$$P_N(x) = \sum_{k=0}^{N} a_k x^k. \tag{4.14}$$

A representation of the procedure is reported, as example, in Figure 4.18, together with the F-test result and the final chosen function used for the multijet reweight.

To objectively select the optimal polynomial degree, thereby balancing model complexity with goodness-of-fit and preventing both underfitting and overfitting, the F-test is employed. The F-test is a statistical tool used to compare the fit quality of two nested models: typically a simpler model (e.g., a polynomial of degree N) and a more complex model (e.g., a polynomial of degree N+1). It quantifies whether the addition of extra parameters in the more complex model leads to a statistically significant improvement in the fit. For two nested models, where a simpler model p has n_p parameters and a more complex model q has n_q parameters (such

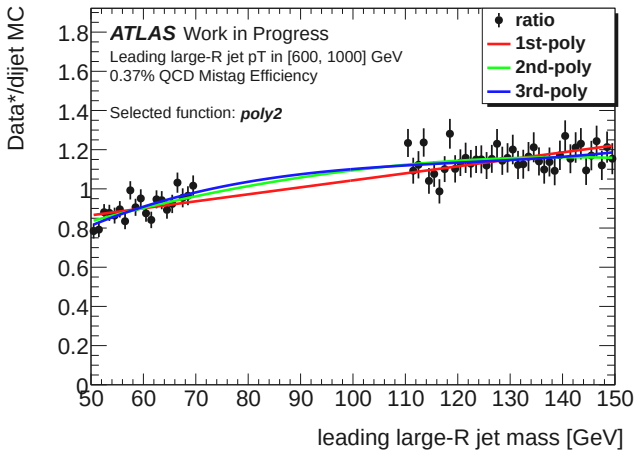


Figure 4.18: Sideband fit of the Data*/MC dijet contribution with different polynomial functions for the GN2X cut corresponding to the 0.37% of mistag efficiency in the p_T interval [600, 1000] GeV. Black points show the bin-by-bin Data*/MC ratio with statistical uncertainties; the three curves are polynomial fits of order 1,2 and 3, used to model the multijet reweighting function. The function choice is based on sequential F-tests with significance threshold $p = 0.05$: F(1 vs 2) $= 1.46 \times 10^{-3}$, F(2 vs 3) $= 1.15 \times 10^{-1}$, so the 2nd-order polynomial is chosen in this configuration.

that $n_q > n_p$), and both models are fitted to n data points (bins), the F-statistic is calculated as:

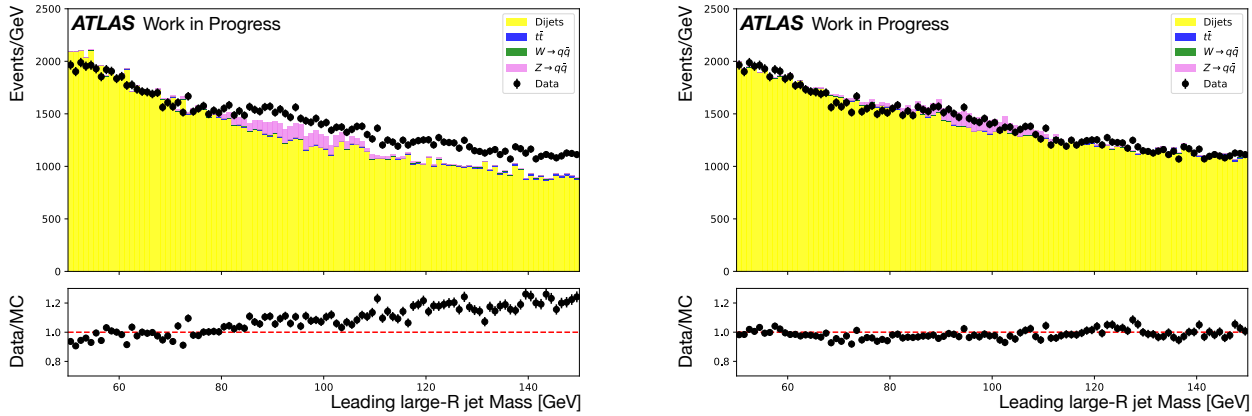
$$F_{p,q} = \frac{\chi_p^2 - \chi_q^2}{n_q - n_p} \Big/ \frac{\chi_q^2}{n - n_q} \tag{4.15}$$

Here, χ_p^2 is the chi-squared value for the simpler model with n_p parameters, and χ_q^2 is the chi-squared value for the more complex model with n_q parameters. The term $(\chi_p^2 - \chi_q^2)$ in the numerator represents the reduction in the chi-squared (improvement in fit quality) achieved by moving from the simpler to the more complex model, and this improvement is normalized by the difference in the number of parameters $(n_q - n_p)$ added.

The denominator, $\chi_q^2/(n - n_q)$, represents the "average" chi-squared per degree of freedom for the more complex model, serving as a baseline for comparison. A large F-value, associated with a small p-value (typically below 0.05), indicates that the more complex model provides a statistically significant improvement in the fit compared to the simpler model, justifying the increased complexity. This F-test is applied iteratively, comparing successive polynomial degrees, to identify the most suitable function that describes the ratio distribution.

Finally, the values of this optimally chosen fitted polynomial function, evaluated at the center of each dijet mass bin, serve as the reweighting factors. These 'pull values' effectively represent the smooth, corrected Data/MC ratio for the QCD background. Each event in the simulated dijet MC sample is then assigned a weight corresponding to the value of this fitted function at its particular dijet mass. By applying these weights, the dijet MC distribution is adjusted bin-by-bin, ensuring its shape and normalization more accurately reflect the QCD background expected in experimental data. This significantly improves the overall data-to-simulation agreement, leading to a more robust background model for subsequent physics measurements.

Distributions before and after the reweighting procedure are reported, for the 70% GN2X working point and for one of the p_T bins, in Figure 4.19.



(a) (b)

Figure 4.19: Leading large-R jet distribution on data and MC samples before (a) and after (b) the multi-jet reweighting procedure showed for events with leading large-R jet p_T in the [500,600] GeV interval after applying GN2X selection corresponding to the 70% signal efficiency. The ratio panel for each plot shows the agreement between data and MC distributions across all the mass fit range, and become closer to one after the reweight is performed.

4.5.2 Spurious Signal Test

After correcting the multi-jet distributions to improve the agreement between data and simulation, the next critical step is to identify the most suitable function to model the QCD background in the S+B fit. This is accomplished using the *spurious signal test*, a procedure specifically designed to quantify the bias introduced by the choice of a background parametrization and ensure that the extracted signal yield is not artificially affected.

The core idea of the spurious signal test is to quantify the amount of artificially introduced signal a given background model would produce when fitting a *background-only* distribution. In other words, even in the absence of true signal events, an imperfect or overly flexible background function may absorb statistical fluctuations in a way that mimics a genuine signal. The magnitude of this artificial signal is then used to guide the selection of the background model that introduces the smallest bias.

In practice, the procedure proceeds as follows:

1. A background-only mass distribution is prepared using the reweighted QCD dijet MC template.

2. An S+B template is constructed, where the signal component corresponds to the $Z(q\bar{q})$ MC distribution after the GN2X tagging requirement as will be used in the final fit on data, and the background component is described by one of the candidate analytic functions.

3. The fit is performed using the S+B template on the background-only distribution. Ideally, the fitted signal strength μ should be zero, but in reality, it can deviate due to mismodelling or statistical fluctuations.

4. The fitted values of μ and its associated uncertainty σ are recorded. The ratio μ/σ serves as a quantitative metric for the spurious signal introduced by that specific background parametrization.

This test is repeated for a wide range of candidate functions, including polynomial forms of different orders (`polyN`) and exponential polynomials (`expolyN`). A function is generally

considered acceptable if it satisfies the criterion $\mu/\sigma < 0.2$, if no function passed this criteria, the threshold is changed to 0.35. If multiple functions pass this requirement, the one with the smallest number of free parameters is chosen in order to minimize the complexity of the final fit and reduce the risk of overfitting. The corresponding μ value is then propagated as a systematic uncertainty on the final $\mu^{\text{post-tag}}$ measurement and, consequently, on the derived scale factors.

The test is performed separately for each p_T interval and for each tagger working point. The candidate functions are defined as:

$$\text{polyN} = \sum_{i=0}^{N} a_i \left(\frac{m}{100\,\text{GeV}}\right)^i, \qquad \text{expolyN} = a_0 \exp\left[\sum_{i=0}^{N} a_i \left(\frac{m}{100\,\text{GeV}}\right)^i\right], \tag{4.16}$$

where m denotes the large-R jet mass and a_i are the free parameters of the fit.

Smoothing the Reweighted QCD Template Running the spurious signal test directly to the reweighted QCD distributions can produce biased results by the presence of significant bin-to-bin statistical fluctuations. To obtain a stable and unbiased background template, the reweighted multi-jet distributions to be fitted are therefore smoothed using a Gaussian Process Regressor (GPR), which is a flexible non-parametric regression technique widely employed in situations where one wishes to model an underlying smooth function from noisy observations [53].

Gaussian Processes (GPs) are in fact probabilistic model for functions which, instead of assuming a specific functional form, such as a polynomial or an exponential, assumes that the value of the function at any set of points follows a multivariate Gaussian distribution. The behavior of the function is controlled by a *kernel*, which specifies how strongly the function values at two different points are correlated through a set of hyper parameters. A smooth kernel, for example, produces functions that vary slowly, while a kernel with a short correlation length allows sharper variations. In Gaussian Process Regression, we treat the unknown "true" background shape as such a function and the observed histogram bins as noisy measurements of it.

Once the GP prior and the kernel are chosen, the method provides the posterior prediction of the function at any point. This prediction is a Gaussian distribution whose mean gives the smoothed value of the histogram and whose variance quantifies the uncertainty of the prediction. Because the kernel enforces correlations across nearby bins, rapid bin-to-bin fluctuations are automatically suppressed, while broader trends in the data are preserved.

In this way, GPR produces a smooth and stable version of the reweighted distribution without imposing a rigid parametric form while keeping essential features of the background. This makes GPR particularly well suited for preparing the templates used in the spurious signal test.

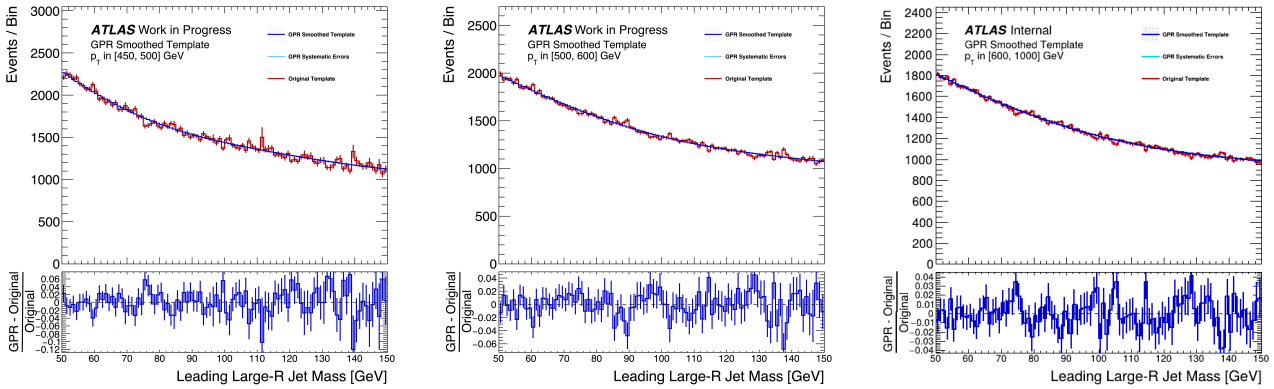
In practice, each reweighted background distribution is smoothed independently for every p_T interval and for each working point. The smoothing is performed using Gaussian Process Regression with a radial basis function (RBF) kernel, which is widely used when one expects the underlying function to be smooth. The RBF kernel is defined as

$$k(x, x') = \sigma_f^2 \exp\left[-\frac{(x - x')^2}{2l^2}\right], \tag{4.17}$$

where σ_f controls the overall variance of the function and l is the characteristic length scale. The length scale determines how quickly the function is allowed to vary: a large l imposes strong correlations between nearby points and therefore produces very smooth functions, while a small l allows the function to change more rapidly. These hyperparameters are determined

from the data by maximizing the marginal likelihood, ensuring that the smoothed curve follows the global shape of the distribution without reproducing bin-to-bin statistical noise.

Conceptually, the Gaussian Process prior defines a probability distribution over an infinite family of possible functions. After observing the reweighted distribution, the GP posterior assigns higher probability to functions that match the data within the noise model, and strongly suppresses functions that oscillate too quickly or contradict the kernel assumptions. The final smoothed template is simply the posterior mean: this can be interpreted as the average of all functions consistent with the kernel properties and the observed distribution, weighted by how well each function fits. Functions that are too irregular or incompatible with the assumed smoothness contribute negligibly, while smooth functions that track the global trend dominate the average.



(a) (b) (c)

Figure 4.20: Smoothed QCD distributions obtained applying the GPR smoothing for the 70% working point of GN2X for each p_T bin used in the calibration.

Figure 4.20 shows an example of the GPR smoothing for the QCD distribution after the 70% GN2X cut application across all the p_T ranges. The upper panel compares the reweighted (unsmoothed) template to the GPR prediction and its associated uncertainty band.

The lower panels display the fractional difference between the GPR prediction and the original template. The residuals fluctuate around zero without any systematic trend, demonstrating that the smoothing does not distort the global shape of the distribution.

Fitting Smoothed and Reweighted QCD Distributions with S+B Template The S+B fit on reweighted and smoothed background only QCD mass distribution is done with the `quickFit` library, which provides a robust flamework for applying statistical models to data. `quickFit` operates using XML workspaces, which serve as structured configuration files. These workspaces allow users to define various components necessary for the fitting procedure in a clear and organized manner. Specifically, the XML workspace includes:

- **Model Definition**, in which the S+B template is outlined using the previous described signal template and the tested poly or expoly function;
- **Data Input**, which in this case consists in the reweighted and smoothed background only multi-jet MC distributions;
- **Parameter Settings**, where initial parameter values and any constrain related to the background model and signal are defined together with the parameters of interest;
- **Fitting Options**, such as the choice of optimization algorithms and other parameters relevant to the fitting process.

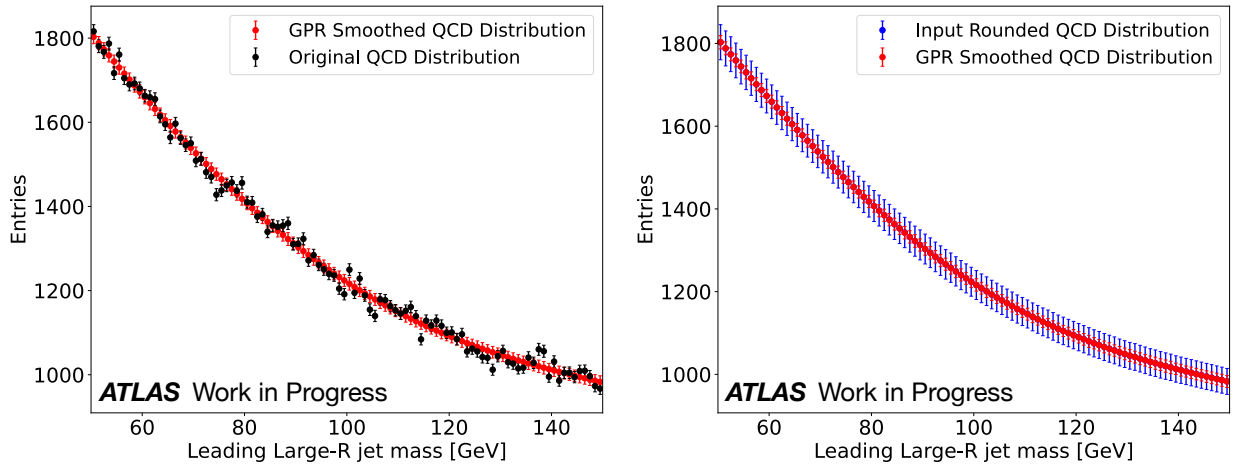


Figure 4.21: Left: The original mass distribution versus the smoothed distribution in the [600, 1000] GeV range at the 70% working point. The smoothed distribution shows reduced statistical fluctuations, minimizing the likelihood of including spurious signals. Right: The comparison between the smoothed and rounded distributions. This plot illustrates the increased Poissonian errors associated with each bin in the rounded distribution, highlighting the substantial impact this has on chi-squared values during fitting.

Once the XML workspace is configured, `quickFit` executes the fit by reading the provided inputs and specifications: the parameter of interest for the fitting is the μ value and the QCD representative function's parameters, and Hessian technique for error estimation and the MINOS algorithm for producing asymmetrical errors on the fitted parameters are used.

During this research, we found that running `quickFit` to fit the mu value on smoothed distributions necessitates a preliminary step to modify the input distributions, making them more Poissonian in nature as the framework is designed for fitting real data distributions. Consequently, an additional manipulation step is required for the reweighted and smoothed QCD mass distributions before executing the fit.

Specifically, each wp-p_T bin distribution is transformed changing the entries to resemble a more data-like structure. This process involves rounding the number of entries (N) for each bin to ensure that the distributions conform to the Poissonian behavior expected in statistical data. Additionally, the associated bin errors are recomputed as $\sqrt{N}$, effectively aligns the error estimate with the poissonian-like expected behavior.

Figure 4.21 provides a visual representation of the error differences the different version of a given input mass distribution. On the left, the comparison between the original mass distribution and its GPR smoothed version in the [600, 1000] GeV leading large-R jet p_T range at the 70% working point of GN2X is reported. The smoothed distribution eliminates statistical fluctuations present in the original data, thereby reducing the risk of introducing spurious signals while fitting. On the right, the comparison is done between the smoothed distribution and its *rounded* counterpart, which represent the final fitted distribution. The comparison highlights the increased error bars resulting from the rounding process, leading to the unphysical chi-squared values obtained when fitting the rounded distributions directly.

To address this issue with the Poissonian errors linked to the smoothed distribution, our final approach involves generating N Poissonianly distributed toys around the rounded distribution. This method effectively circumvents the mismatch between the behavior of the fitted points and their associated errors. By simulating multiple toy distributions that follow Poisson statistics, we ensure that the fitting process accurately reflects the underlying statistical nature of the data, thus enhancing the reliability of the fitting results. The complete procedure, with the

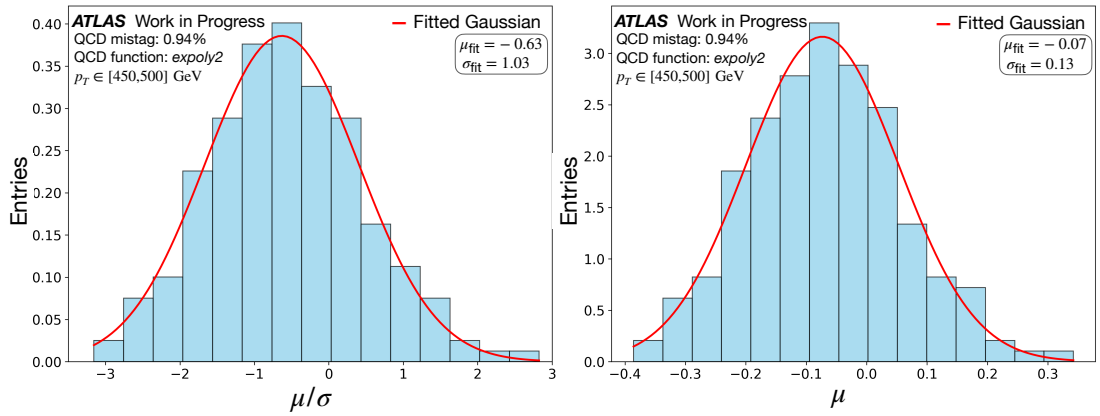


Figure 4.22: μ/σ and μ only distributions obtained running the S+B fit on 200 toys poissonianly generated around the smoothed and rounded QCD Monte Carlo distribution for the 70% GN2X working point, p_T in [450,500] GeV and expoly2 as QCD function. Both the distributions are fitted with gaussian functions and the fitted mean and standard deviation values are used as metric and systematic uncertainties.

final results obtained with it, is reported in the next paragraph.

Spurious Signal Test on Toys To compute the final results of the spurious signal test, a total of 200 toys were generated around the rounded distribution, employing Poissonian statistics to closely align with the expected behavior of the data.

Each generated toy underwent a signal-plus-background fit with the background template described time by time with different functions: polynomial of degree 3 (poly3), exponential polynomial of degree 2 (expoly2), exponential polynomial of degree 3 (expoly3), and exponential polynomial of degree 4 (expoly4). The `quickFit` output values, μ and σ, were then stored for each configuration (WP, p_T range, QCD function).

The μ and σ values were then systematically organized into distributions tailored for each specific WP, p_T and background function configuration. To analyze these distributions, both the μ values and the μ/σ ratios were defined and evaluated by fitting with a Gaussian, as depicted in figure 4.22. Mean and standard deviation for each distribution.

The mean value obtained from the Gaussian fit of the μ/σ distribution served as a critical metric for evaluating whether the fitted background function met the acceptance criteria (μ / σ < 0.2), while the mean value of the μ only distribution is used as representative of the systematic uncertainty related to the choice of each modelling function.

Tables moved into the Appendix for a better clarity of the text, summarize the mean value and standard deviation of the μ and μ/σ distributions for each tested function and configuration used for selecting the optimal function to describe the background in the final signal-plus-background (S+B) fit on the data distribution.

There, functions highlighted in green indicate those that successfully pass the stricter $\mu/\sigma <$ 0.2 selection. Conversely, if a function was ultimately chosen because it only met the relaxed $\mu/\sigma < 0.35$ criterion (meaning no functions passed the 0.2 cut), it is highlighted in yellow.

While the optimal modeling function determination was comprehensively performed for each working point across all defined p_T ranges, we now specifically highlight the selections for

the 70% signal efficiency working point, corresponding to the 0.94% QCD efficiency, as this particular working point is the one that will be used within the VHbb analysis in the next chapter.

Summarizing, for the 70% WP the chosen background modelling functions, with associated systematic uncertainties, are:

Table 4.7: Summary of Spurious Test Results for 70% Working Point

p_T Range	[450, 500] GeV	[500, 600] GeV	[600, 1000] GeV
QCD Function	**expoly3**	**expoly4**	**expoly4**
μ/σ	-0.06	-0.23	-0.04
μ	-0.01	-0.04	0
σ	0.13	0.16	0.16

4.5.3 Results: $\mu^{post-tag}$ values

With the optimal background models identified for the 70% working point in each p_T interval, the final step before defining the Scale Factors is the extraction of the $\mu^{\text{post-tag}}$ values. This is achieved by fitting the data leading large-R jet distribution using the full S+B template. The procedure follows exactly the same fitting framework and methodology previously validated through the spurious-signal tests, ensuring robustness and internal consistency.

The signal (S) component of the template, representing the expected $Z(b\bar{b})$ signature, is derived from $Z(q\bar{q})$ simulations after the GN2X cut application corresponding to the 70% signal efficiency, providing the precise shape of the expected signal mass distribution. The background (B) component for each p_T bin is defined using the specific polynomial functions (expoly3 or expoly4) that were selected based on the spurious test analysis for the 70% working point, as detailed in Table 4.7.

The S+B fit then determines the overall normalization factor, $\mu^{\text{post-tag}}$, that best scales this signal template to match the observed data.

The results of these S+B fits on the large-R jet mass distribution for each p_T interval are graphically presented in Figure 4.23. Each figure shows the observed data (black points with error bars) alongside the various components of the fit:

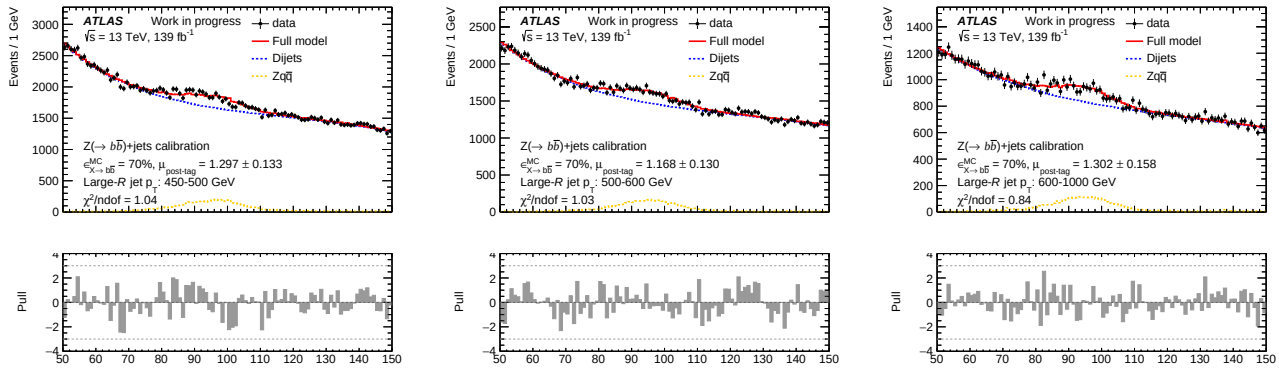
- The `Full model` (solid red line) represent the combined S+B template, demonstrating the overall fit to the data;
- The `Dijets` (blue dotted line) illustrates shape of the expoly function used for representing the QCD component in the fit;
- The `Zqq` (yellow dotted line) represents the $Z(b\bar{b})$ signal component fitted in data.

Each plot contains also the statistical uncertainties associated to each fit, together with the χ^2/ndof quantity (chi-squared per number of degrees of freedom), which is close to 1 for each p_T bin and indicates that the used template describes well the data without underfitting or overfitting.

The pull distributions, shown below each mass distribution, provide a more granular view of the fit quality. Each bar in the pull plot represents the difference between the observed data and the fitted model, divided by the total uncertainty in that bin, and the distributions obtained across all the p_T bins demonstrates that the residuals are statistically consistent with the uncertainties, and there are no significant systematic mis-modeling effects or large, localized deviations, thus reinforcing the reliability of the S+B fits.

p_T [GeV]	**450 - 500**	**500 - 600**	**600 - 1000**
$\mu^{\text{post-tag}}$	1.297	1.168	1.302
Statistical Uncertainty (+)	0.133	0.130	0.158
Statistical Uncertainty (-)	-0.130	-0.131	-0.158

Table 4.8: Summary of fitted $\mu^{\text{post-tag}}$ values along with their statistical uncertainties.



(a) (b) (c)

Figure 4.23: S+B fit of data leading large-R jet mass distribution for each p_T interval. Signal template is extracted using $Z(q\bar{q})$ events passing the GN2X 70% working point, the background template is obtained through the spurious signal test and is expoly3 (a), expoly4 (b) and expoly4 (c) function.

p_T [GeV]	[450 - 500]	[500 - 600]	[600 - 1000]
$\mu^{\textbf{post-tag}}$	1.297	1.168	1.302
Statistical Error	± 0.133	± 0.130	± 0.158
Systematics			
$Z(b\bar{b})$+jets Modelling	+ 0.120/-0.214	+0.028 / -0.026	+ 0.026 / -0.030
W Modelling	+ 0.003/-0.007	+0.005 / -0.005	+ 0.0042 / -0.006
$t\bar{t}$ Modelling	+ 0.028/-0.032	+0.017 / -0.017	+ 0.010 / -0.012
Background Modelling	± 0.056	± 0.100	± 0.070
JMS	+ 0.019/-0.011	+0.011 / -0.003	+ 0.021 / -0.007
JMR	± + 0.148/-0.220	+0.088 / -0.105	+ 0.065 / -0.082
JES	+0.072 / -0.071	+0.063 / -0.061	+0.031 / -0.029

Table 4.9: Summary of $\mu^{\text{post-tag}}$ values along with their statistical and systematic uncertainties.

Beyond statistical uncertainties, a comprehensive assessment of the $\mu^{\text{post-tag}}$ value requires evaluating potential systematic uncertainties.

Systematic uncertainties related to jets play a critical role when dealing with hadronic final states [54]. The Jet Mass Scale (JMS) captures the uncertainties in the calibration of jet mass measurements, which can arise from variations in detector response and the influence of underlying events. The Jet Mass Resolution (JMR) assesses the precision with which the jet mass can be reconstructed, taking into account factors such as detector resolution and the effects of additional pile-up collisions, which may distort the measured mass. Finally, the Jet Energy Scale (JES) includes systematic uncertainties in the overall energy calibration of jets, reflecting the accuracy of energy measurement and corrections for differing particle types and their interactions in the detector. The values of these uncertainties were derived from previous calibration studies of the Xbb tagger, which used the same Z-based procedure, serving as a

crucial reference for this analysis.

These uncertainties contribution, together with the Z, W and $t\bar{t}$ modelling ones, are reported in Table 4.9.

The background modelling systematic uncertainties are defined as the quadrature sum of the results of the following studies:

1. **Spurious signal test**: measuring the contribution from background function choice by fitting a background only sample with a S+B template;

2. **Signal injection test**: measuring additional biases introduced by the QCD chosen modelling function in presence of the signal;

3. Stability of the fit across different **mass ranges**: investigating the sensitivity of the fitted $\mu^{\text{post-tag}}$ value to variations in the mass range used for the fit, exploring the impact of using smaller, larger, or shifted mass windows.

4.5.4 Background function related uncertainties

The systematic contribution arising from the choice of the QCD function for each p_T bin is quantified as spurious bias included into the S+B fit for each chosen background function, and μ values reported in Table 4.7 are directly considered as systematic contribution.

An additional estimation of the signal extraction efficiency and potential bias is provided by the signal injection test.

The signal injection test offers an additional estimation of signal extraction efficiency and potential bias, serving as a critical methodology to evaluate how accurately the fitting framework recovers a known signal in the presence of background. This test involves performing a signal-plus-background (S+B) fit on simulated distributions, where the true signal strength is known (tipically set to unity). Using the background function selected in the spurious test, the difference between the fitted μ value and the true (injected) μ value (which is 1) allows for quantifying any potential bias from the fit.

The process utilizes the same toy-based approach and as the spurious test and the same fitting framework as the spurious tests and the final S+B fit.

It constructs a "well-known" distribution by combining a simulated QCD mass distribution (reweighted, smoothed, and rounded) with a precisely defined amount of the $Z(q\bar{q})$ signal template. By design, the true signal strength injected into this pseudo-data is explicitly set to unity ($\mu_{\text{true}} = 1$). Subsequently, the S+B fit is conducted on a set of Poisson-distributed toy samples.

An unbiased and efficient fitting procedure should ideally recover a fitted $\mu^{\text{post-tag}}$ value of 1 when applied to such a precisely defined distribution. To assign the fitted $\mu^{\text{post-tag}}$ for each configuration in the signal injection test, a set of toy S+B MC distributions are generated and fitted with the S+B template. The resulting distribution of $\mu^{\text{post-tag}}$ values from these individual fits undergoes Gaussian fitting, with the mean of this Gaussian distribution taken as the final fitted $\mu^{\text{post-tag}}$ value for that configuration.

Any absolute difference observed between this mean fitted $\mu^{\text{post-tag}}$ value and the true injected value of 1 quantifies a systematic bias or inefficiency within the signal extraction method. This specific difference, denoted as $|\mu^{\text{post-tag}}_{\text{fitted}} - 1|$, is then assigned as a systematic uncertainty for each respective p_T interval, providing a direct measure of the accuracy and potential bias in the signal strength recovery.

A summary of the systematics related to the background modelling function, together with the statistical uncertainty for the final $\mu^{post-tag}$ fit, used for reference, is reported in Table 4.10:

p_T [GeV]	**450 - 500**	**500 - 600**	**600 - 1000**
QCD Function	expoly3	expoly4	expoly4
Stat σ	0.133	0.130	0.158
Spurious μ	-0.01	-0.04	0.0
Injected $\lvert\mu - 1\rvert$	0.05	0.03	0.01

Table 4.10: Statistical uncertainties associated with the background modelling function used in the final S+B fit to data for the extraction of $\mu^{\text{post-tag}}$. For each p_T range, the QCD function is determined through the spurious signal test described above. The quoted spurious uncertainties correspond to the fitted μ values obtained from the toy-based procedure, and are summarized in Table 4.7. The signal-injection systematic uncertainties, evaluated in a similar way as the difference between the injected value ($\mu = 1$) and the fitted one, are also reported.

A clear observation from the table is that the statistical uncertainty related to the final fit is consistently and substantially larger than both the systematic uncertainty from the spurious and the signal injection test across all p_T bins. For instance, in the $[450, 500]$ GeV range, the statistical uncertainty is 0.13, while the spurious μ is -0.01 and the injected $|\mu - 1|$ is 0.05. This demonstrates that these two systematic effects are currently subdominant and do not significantly impact the overall uncertainty budget of the $\mu^{\text{post-tag}}$ measurement, highlighting the robustness of the statistical precision achieved in the signal extraction process.

4.5.5 Mass fit range related uncertainties

Beyond the systematic contributions from the background function choice and signal injection, a crucial aspect of validating the $\mu^{\text{post-tag}}$ measurement is to assess its stability across different mass ranges. This systematic uncertainty quantifies any potential bias or sensitivity that might arise from the specific selection of the mass window over which the signal-plus-background (S+B) fit is performed.

The stability of the fit with respect to the chosen mass range is investigated to verify that the extracted results are not strongly dependent on the specific interval used in the fit. In the nominal configuration, the fit is performed in the mass range $[50, 150]$ GeV.

The definition of the fitting range is relevant because the continuous QCD background is described using simplified empirical functions, such as expoly3 or expoly4. While these functions provide a flexible parametrization of the smoothly falling background spectrum, their ability to reproduce the underlying distribution may depend on the mass interval considered. For this reason, variations of the fit range are studied to ensure that the fitted signal yield and background description remain stable.

In addition, extending the range significantly beyond the nominal region, particularly towards higher masses, may introduce contributions from other processes, such as $t\bar{t}$ production, whose mass distributions are not described by the chosen background parametrization. The nominal range is therefore selected to balance background modelling accuracy and robustness of the fit.

To thoroughly investigate the impact of these factors, nearby ranges were systematically tested: $[50, 140]$ GeV, $[50, 155]$ GeV, and $[55, 150]$ GeV. These variations were intentionally kept small, especially at the high-mass end, to avoid inadvertently including a large and complex $t\bar{t}$ contribution that the current background models are not specifically designed to handle. This approach allows for a focused assessment of the $\mu^{\text{post-tag}}$ measurement's sensitivity to minor changes in the mass window, without introducing new, unmodeled physics effects. For these tests, the background function determined in the standard fit was used, and the largest difference in $\mu^{\text{post-tag}}$ values compared to the nominal fit is then taken as a systematic uncertainty

for each p_T bin.

Table 4.11: Fit Range Stability for p_T in [450, 500] GeV and $\mu^{post-tag} = 1.297 \pm 0.133$

Fit Range [GeV]	**Fitted $\mu^{\text{post-tag}}$**	$\Delta\mu$
[50, 140]	1.29794	-0.0009
[50, 155]	1.32490	-0.02786
[55, 150]	1.26786	0.02918

Table 4.12: Fit Range Stability for p_T in [500, 600] GeV. Nominal $\mu^{\text{post-tag}} = 1.168 \pm 0.13$.

Fit Range [GeV]	**Fitted $\mu^{\text{post-tag}}$**	$\Delta\mu$
[50, 140]	1.13181	0.03637
[50, 155]	1.16035	0.00783
[55, 150]	1.26918	-0.10100

Table 4.13: Fit Range Stability for p_T in [600, 1000] GeV. Nominal $\mu^{\text{post-tag}} = 1.302 \pm 0.158$.

Fit Range [GeV]	**Fitted $\mu^{\text{post-tag}}$**	$\Delta\mu$
[50, 140]	1.39143	-0.08979
[50, 155]	1.31757	-0.01593
[55, 150]	1.37862	-0.07698

Using the $\Delta\mu$ values obtained for each configuration, the largest variation with respect to the fitted signal strength in the nominal fit range $[50, 150]$ GeV is taken as a systematic contribution, normalized to the nominal value. The resulting uncertainties for each p_T bin are reported in Table 4.14.

Table 4.14: Relative Systematic Uncertainty from Fit Range Stability

p_T range [GeV]	[450,500]	[500,600]	[600,1000]
$\Delta\mu/\mu$	0.0225	-0.0865	-0.069

The values demonstrate that while the fit generally shows good stability, the choice of fitting range introduces a non-negligible systematic uncertainty, particularly for the higher p_T bins. These relative uncertainties will be propagated into the overall systematic uncertainty budget for the $\mu^{\text{post-tag}}$ measurement.

4.5.6 Scale Factor Extraction

As anticipated, the final Scale Factors are defined as the ratio between the post-tag and pre-tag signal strengths:

$$\text{SF} = \frac{\mu^{\text{post-tag}}}{\mu^{\text{pre-tag}}}, \tag{4.18}$$

where the pre- and post- tag signal strengths are measured on leptonic and hadronic Z-boson final states respectively as discussed before.

Because the SF is constructed as a ratio, any source of systematic uncertainty affecting both $Z(l^+l^-)$+jets and $Z(b\bar{b})$+jets measurement will cancel out in the ratio to first order. Both measurements rely on the same theoretical systematics (PDF and scale variations) and the same

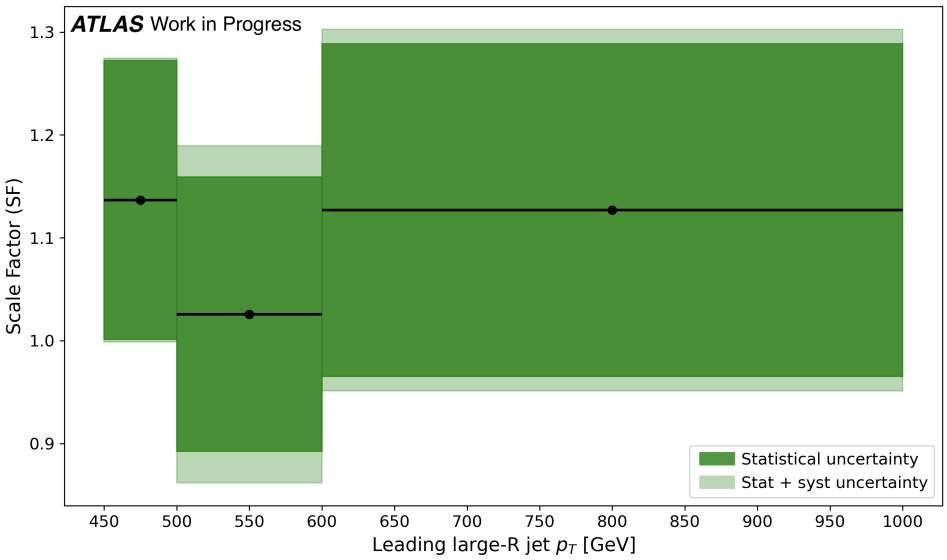


Figure 4.24: Scale Factor values for GN2X 70% signal efficiency as a function of the leading large-R jet p_T.

luminosity, so the related uncertainties have not to be included in the SF one. Also, tables 4.3 and 4.5 show that the same MC generators were used to produce the Z, W, and $t\bar{t}$ samples. As a consequence, the matrix element modelling uncertainties largely cancel when considering ratios between the $Z \to b\bar{b}$ and $Z \to \ell^+\ell^-$ processes. In general, however, parton shower modelling uncertainties would not be expected to cancel completely, since the shower configurations are not identical for all samples. As this study is still preliminary and the corresponding parton shower systematic variations have not yet been evaluated, these effects are not included in the current uncertainty estimate.

The total SF uncertainty therefore includes the statistical errors on both signal strength measurements, the systematic components from the pre-tag measurement related to the di-boson modelling, and all post-tag systematics described in the previous sections related to the background modelling, including the spurious-signal results, the signal-injection test, and variations of the mass-fit range.

Following this procedure, the Scale Factor values for each p_T interval corresponding to the flat-mass QCD 70% working point of the GN2X discriminant are shown in Figure 4.24.

The corresponding values are reported in Table 4.15.

Large-R jet p_T [GeV]	**450 - 500**	**500 - 600**	**600 - 1000**
Scale Factor	1.1368	1.0256	1.1550
Statistical Uncertainty	0.1357	0.1334	0.1617
Systematic Uncertainty	0.0246	0.0953	0.0690
Total Uncertainty	0.1379	0.1639	0.1758

Table 4.15: Scale Factor values for the 70% WP.

These Scale Factors quantify the residual differences between data and simulation in the GN2X tagger performance after applying the 70% signal-efficiency working point. As shown in Fig. 4.24 and summarized in Table 4.15, the GN2X tagger exhibits SF values close to unity across the full large-R jet p_T range considered, indicating an overall good modelling of the tagger response in Monte Carlo. The uncertainty budget is dominated by the statistical precision of the extracted signal strengths, with systematic contributions remaining subdominant.

In the next chapter, the impact of this new tagging strategy on the $VH(b\bar{b})$ boosted analysis will be assessed. In particular, the improvements in signal sensitivity and background rejection achieved through the GN2X tagger, once calibrated and validated as presented here, will be quantified and compared to the performance of previous tagging techniques, demonstrating the full physics potential unlocked by the calibration delivered in this chapter.

Chapter 5

Higgs decay to a pair of $b-$ quarks in association with a vector boson in the high-p_T regime

The discovery of a heavy scalar particle with a mass of approximately 125 GeV by the ATLAS and CMS collaborations at the Large Hadron Collider in 2012 marked a milestone in particle physics [1], [2]. Subsequent measurements of its properties confirmed that this particle is indeed the Higgs boson predicted by the Brout-Englert-Higgs (BEH) mechanism, responsible for the spontaneous breaking of electroweak symmetry and the generation of masses for the W and Z gauge bosons. Further observations of Higgs boson decays to fermions, including decays to τ-leptons [55] [56], bottom quarks [57] [58], and production in association with top quarks [59] [60], have provided compelling evidence that Yukawa interactions operate as predicted within the Standard Model (SM). To date, all measured Higgs couplings remain compatible with SM expectations.

Among all Higgs decay channels, the $H \to b\bar{b}$ decay to a pair of bottom quarks plays a central role. This is mainly due to its large branching fraction, which is predicted to be approximately 58% for a Higgs boson mass of 125 GeV. As a result, the bottom-quark Yukawa coupling has a dominant influence on the total Higgs width. Precise measurements of this coupling therefore provide one of the most sensitive probes of potential deviations from the SM and are essential for constraining scenarios of physics beyond the Standard Model.

In this context, the associated production of the Higgs boson with a vector boson (VH, where $V = W, Z$) constitutes a particularly advantageous production mechanism for studying the $H \to b\bar{b}$ decay. The presence of a high-momentum vector boson in the final state offers a clean experimental signature that significantly enhances the suppression of the overwhelming QCD multijet background. Moreover, the leptonic decays of the accompanying W or Z boson provide additional triggers and powerful discrimination against background processes. For these reasons, the VH topology has long been, and remains, one of the most sensitive channels for the observation and precise measurement of the Higgs-bottom coupling.

In addition, differential measurements of Higgs boson production as a function of its transverse momentum constitute a powerful tool to explore the dynamics of the Higgs sector, as deviations from SM predictions are expected to be more pronounced at high transverse momenta, where new physics effects may enhance the high-p_T region of the Higgs boson spectrum. In this boosted regime, Higgs-strahlung emerges as the dominant production mechanism, where still the Higgs boson predominantly decays to $b\bar{b}$. Efficient reconstruction and identification of boosted Higgs candidates are therefore essential for precision measurements in this phase space.

A crucial aspect of such analyses is the ability to reliably identify bottom-quark jets across a wide kinematic range. In the previous chapter, the calibration strategy for the GN2X boosted

$b\bar{b}$ tagger was presented, and as this tagger is specifically designed to identify Higgs decays to $b\bar{b}$ in the boosted regime, ensuring that its performance is accurately calibrated in data is essential for its use within this analysis.

Incorporating the calibrated GN2X tagger into the VH, $H \to b\bar{b}$ analysis significantly enhances the ability to isolate signal-like jet topologies, particularly at high p_T, resulting in an overall increased sensitivity of the measurement.

In the following sections, the full physics analysis strategy is presented, illustrating in detail how the calibrated GN2X tagger is incorporated into the signal event selection, how it affects the categorization of events in the boosted regime and how its inclusion ultimately strengthens the measurement of the VH production cross-section and the Higgs-bottom coupling.

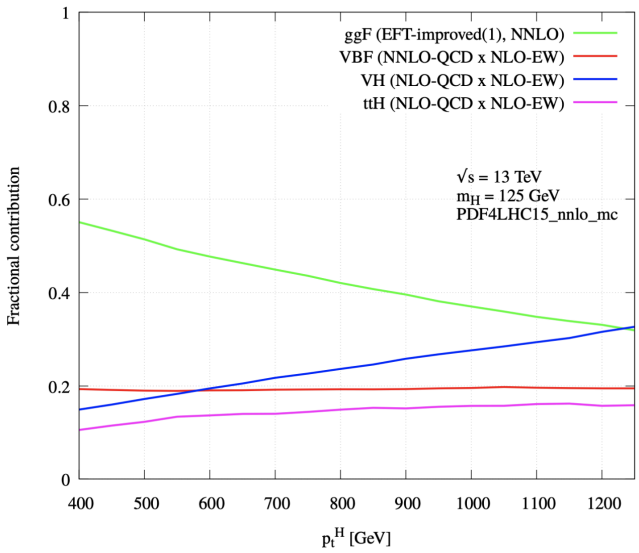


Figure 5.1: Higgs boson production fractions as a function of its transverse momentum p_T^H [61]. Gluon-gluon fusion dominates across most of the p_T^H spectrum, while at higher p_T^H values, production in association with a vector boson becomes comparable.

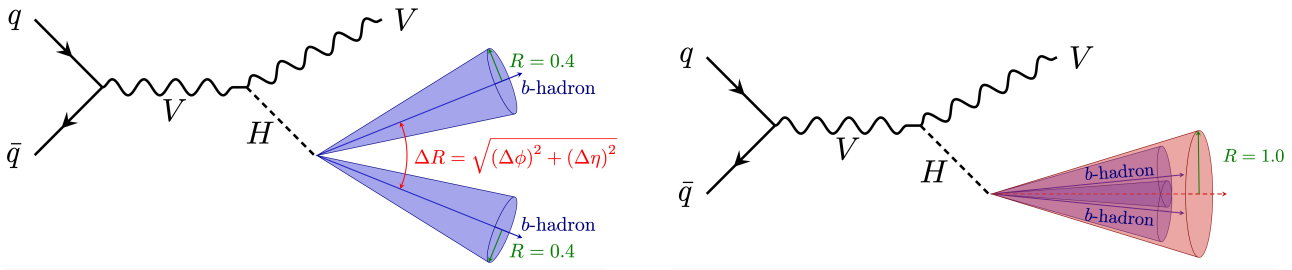
5.1 Boosted $VH, H \to b\bar{b}$ Analysis Strategy

As explained in more detail in Section 1.3, the dominant Higgs production mechanisms at the LHC are gluon-gluon fusion (ggF), vector boson fusion (VBF), vector boson associated production (VH) and top quark associated production (ttH). Figure 5.1 illustrates how the different Higgs production modes contribute variably as the Higgs transverse momentum increases. While ggF is the dominant production mechanism across a wide range of p_T^H, its prevalence poses significant experimental challenges due to the overwhelming QCD background it generates. This complexity makes detailed studies of Higgs final states via the ggF production channel less feasible for precise measurements.

In contrast, the VH production mode, although subdominant compared to ggF, plays a particularly important role in experimental analyses targeting the boosted regime, where the production fraction become comparable as the ggF one. In this topology, the presence of the vector boson provides clean experimental signatures thanks to its leptonic decays.

As the angular separation of decay products depends on the decaying particle p_T, also the analysis is structured into two distinct topologies: resolved and boosted, depending on the transverse momentum of the decaying particle itself. In the resolved topology, the low-p_T Higgs boson decays into two $b-$quarks which can be reconstructed as two distinct small-R jets. Conversely, at high-p_T values, the Higgs boson is boosted enough to reconstruct the in $b\bar{b}$ final

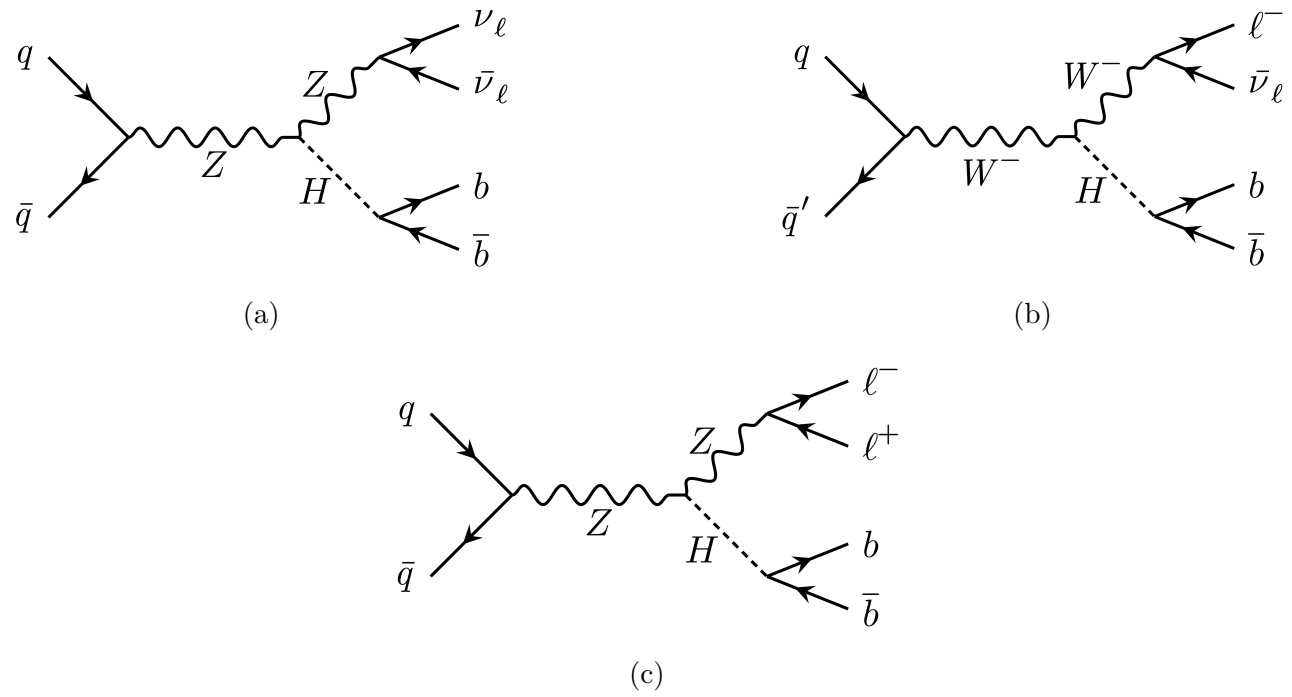
state as a single large-R jet, which can be then tagged using the GN2X tagger described and calibrated in the previous Chapters.



(a) (b)

Figure 5.2: $VH, H \to b\bar{b}$ resolved (a) and boosted (b) topologies [20].

Feynman diagrams illustrating both categories are provided in Figure 5.2, highlighting the distinct reconstruction approach involved in each topology: two distinct small-radius jet with R = 0.4 at low p_T, while as the p_T increases (typically for p_T >250 GeV) the more collimated decay products are reconstructed with a single large radius jet with R = 1.0. As during my PhD I focused on calibrating the GN2X tagger employed for boosted topologies, the analysis topology presented here is the boosted one.



(a) (b) (c)

Figure 5.3: Leading-order Feynman diagrams of the VH production followed by a $H \to b\bar{b}$ decay for the (a) 0-lepton, (b) 1-lepton, and (c) 2-lepton channels ($l = e, \mu$).

For each topology the analysis is thus divided into three orthogonal channels according to the number of reconstructed charged leptons in the final state, specifically electrons and muons, since tau leptons decay rapidly before reaching the ATLAS detector. Each channel targets a distinct decay mode of the vector boson, summarized in Figure 5.3.

More specifically, the three channels in which the analysis is divided are:

- **0-Lepton Channel:** corresponds to $Z \to \nu\bar{\nu}$ decays, characterized by a significant amount of missing transverse energy E_T^{miss} and the absence of any detected leptons.
- **1-Lepton Channel:** involving $W \to l\nu$ decays and requires the identification of one charged electron or muon, along with considerable E_T^{miss}. Also events that meet the 0-lepton criteria and feature at least one identified $\tau-$jet are re-classified as part of the 1-lepton channel (as likely indicative of a WH decay with $W \to \tau\nu$).
- **2-Lepton Channel:** targeting the $Z \to l^+l^-$ decays selected requiring two same-flavour electrons or muons with an invariant mass consistent with the Z boson one. Opposite charges are required when muons are in the final state while such criterion is not applied for electrons due to the larger charge mis-identification.

The 0-lepton channel has the largest branching fraction among the three leptonic channels, but it faces challenges related to the reconstruction of missing transverse energy. In contrast, the 1-lepton channel strikes a good balance between a substantial event rate and a manageable background level. This background is further minimized in the 2-lepton channels, which represent the configuration with the lowest background contamination, although they also yield the smallest event count due to the leptonic branching fraction of the Z boson.

The 0- and 1-lepton channels utilize reconstructed missing transverse energy to infer the presence of neutrinos in $ZH \to \nu\nu b\bar{b}$ and $WH \to l\nu b\bar{b}$ events, based on momentum imbalances in the transverse plane. In the 0-lepton channel, the vector boson transverse momentum p_T^V is reconstructed as the missing transverse energy. In the 1-lepton channel, it is represented by the sum of the missing transverse energy E_T^{miss} and the momentum of the charged lepton. In the 2-lepton channel, p_T^V corresponds to the transverse momentum of the two-lepton system.

Despite the selection of leptonic final states, several substantial background contributions remain. Specifically, the background processes to be taken into account when targeting $VH, H \to b\bar{b}$ events can be divided in six categories: $t\bar{t}$, single top, W + jets, Z + jets, di-boson and multi-jet events. Those contributions can be further split into reducible and irreducible backgrounds, depending if they are resonant or not, and summarized in Figure 5.4.

Reducible backgrounds: Reducible backgrounds correspond to events in which a reconstructed $b\bar{b}$ final state does not originate from the signal $VH(b\bar{b})$ process. These backgrounds are non-resonant and can be suppressed since their invariant mass distribution m_{bb} does not peak at the Higgs boson mass and their kinematic properties, such as $\Delta R(b, \bar{b})$ and p_T, do not resemble those expected from the two-body decay $H \to b\bar{b}$.

Representative processes include top-quark pair production, single-top production, and multijet events that pass the object selection due to instrumental limitations. In particular, $t\bar{t}$ production naturally yields two b quarks in the final state, making it a significant source of background when both b-jets are correctly identified. As a result, $t\bar{t}$ events constitute the dominant background contribution in the single-lepton channel. Their impact can be reduced by exploiting additional event information, such as activity outside the candidate large-R jet, which is included among the inputs of the machine learning models used to discriminate signal from background.

Irreducible backgrounds: Irreducible backgrounds arise from processes that produce the same final-state signature as the signal and therefore cannot be distinguished through simple kinematic selections. In the context of the $VH(b\bar{b})$ search, these backgrounds correspond primarily to the associated production of a vector boson with a $b\bar{b}$ pair ($V + b\bar{b}$, $V = Z, W$), where the two b-jets originate from QCD radiation rather than from the decay of a Higgs boson and are mis-classified by the tagger.

Because these processes yield the same visible final state as the signal, their kinematic properties can closely resemble those expected from $H \to b\bar{b}$ decays (as assumed in the GN2X signal calibration procedure). As a consequence, they represent an intrinsic limitation to the analysis sensitivity and cannot be fully removed through event selection alone. Instead, their contribution is controlled through precise background modelling and by exploiting multivariate techniques that combine several event-level features. Other irreducible backgrounds is di-boson production, particularly WZ and ZZ production where the Z boson decays to $b\bar{b}$.

Before presenting the details of the boosted $VH, H \to b\bar{b}$ analysis on which I worked during my PhD, the following section reviews the current state-of-the-art of the $VH, H \to b\bar{b}$ analysis in order to provide context for the work presented, introducing also the main concepts which will be used.

5.2 State-of-the-art: The ATLAS $VH, H \to b\bar{b}/c\bar{c}$ Run 2 Analysis

The most recent study of Higgs boson decays to heavy quarks in the associated production VH mode was published by the ATLAS Collaboration in 2024 [62]. This work presented the results $VH, H \to b\bar{b}/c\bar{c}$, which consists of a reanalysis of the full Run 2 dataset using the updated ATLAS recommendations, including the DL1r flavour-tagging algorithm for $b\bar{b}$ jets selection. The measurement is based on an integrated luminosity of $140\,\text{fb}^{-1}$ of pp collisions at $\sqrt{s} = 13\,\text{TeV}$, and provides both the most precise determination of the $H \to b\bar{b}$ decays to date and the most sensitive direct search for $H \to c\bar{c}$ decays.

In this analysis, a combined study of $VH, H \to b\bar{b}$ (in both the resolved and boosted regimes) and $VH, H \to c\bar{c}$ was performed within a common analysis framework, designed to maximize the sensitivity through a simultaneous treatment of the two decay modes. This approach is well motivated by the strong similarities between the $VH, H \to b\bar{b}$ and $VH, H \to c\bar{c}$ final states, in particular the presence of common dominant background processes, namely V+jets and $t\bar{t}$ production. A coherent modeling of these backgrounds enables a simultaneous extraction of the coupling modifiers κ_b and κ_c, which quantify possible deviations of the Higgs boson couplings to bottom and charm quarks from their Standard Model values, with shared constraints, leading to improved overall precision.

5.2.1 $VH, H \to b\bar{b}/c\bar{c}$ Analysis Strategy

The last published $VH, H \to b\bar{b}/c\bar{c}$ analysis, as the one presented in this thesis, was also divided in leptonic channels, targeting the different vector boson decay modes: 0-lepton for $Z \to \nu\nu$ events, 1-lepton for $W \to l\nu$ decays, and 2-lepton for $Z \to ll$ ones, with l being always an electron or a muon. Also here the reconstructed events containing a τ lepton in the 0-lepton channel were moved into the 1-lepton one. As illustrated in Figure 5.5, the $VH, H \to b\bar{b}$ analysis is divided into resolved and boosted regimes, while $VH, H \to c\bar{c}$ processes are considered over the full p_T spectrum. The resolved and boosted $VH, H \to b\bar{b}$ analyses are separated by a threshold on the vector-boson transverse momentum of $p_T > 400\,\text{GeV}$. Below this value, corresponding to the resolved regime, the Higgs boson decay is reconstructed from two small-radius jets. Above the threshold, in the boosted regime, the Higgs candidate is reconstructed as a single large-R jet with at least two associated subjets. The $VH, H \to c\bar{c}$ category is defined by events that fail the b-tagging requirements but satisfy the corresponding c-tagging criteria.

The analysis treats the $VH, H \to b\bar{b}$ and $VH, H \to c\bar{c}$ processes as the primary signal

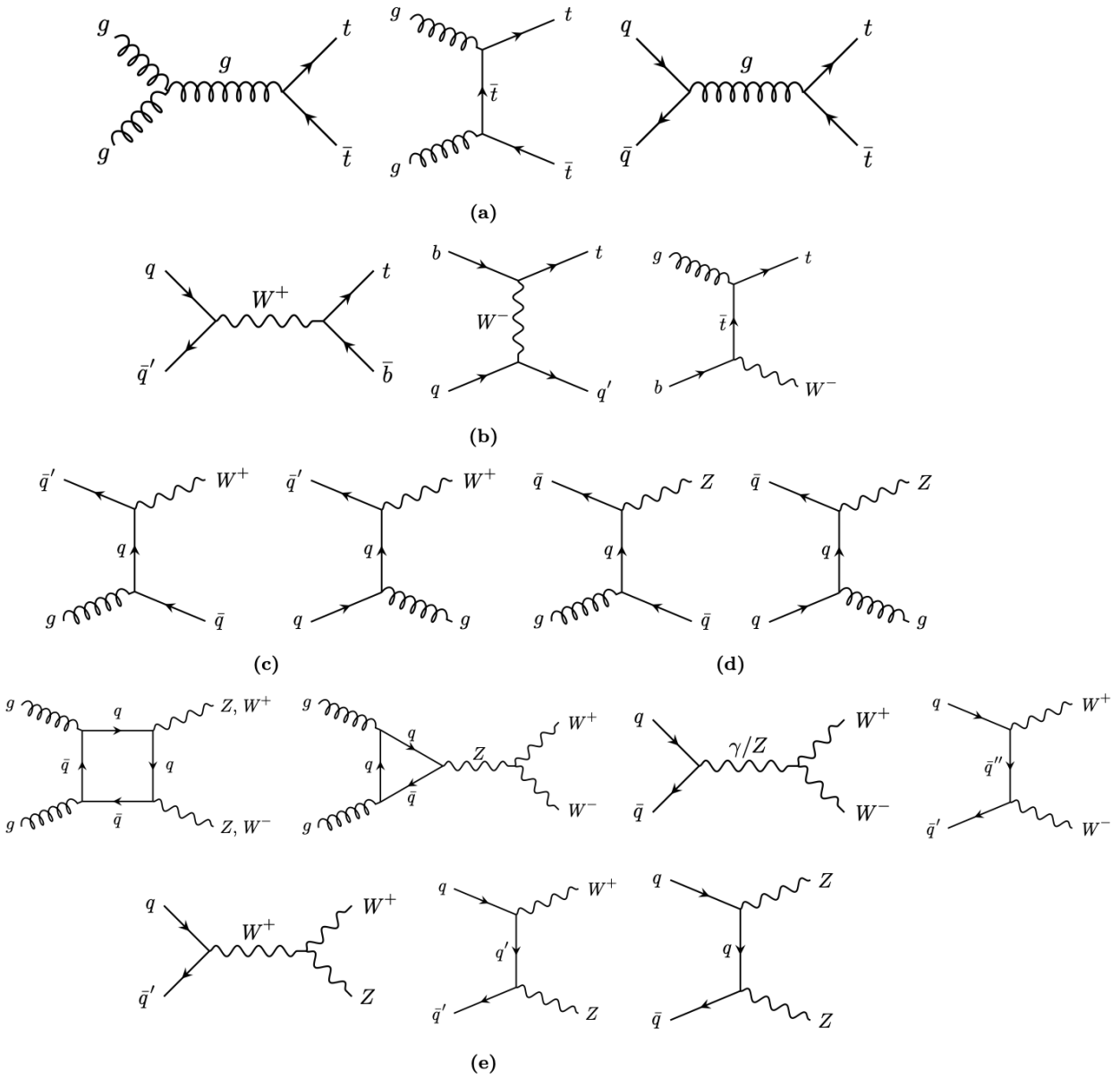


Figure 5.4: Main $VH, H \to b\bar{b}$ background processes: respectively (a) gluon and quark induced $t\bar{t}$ events, (b) s-channel, t-channel and Wt single top events, (c) gluon and quark induced W+jets events, (d) gluon and quark induced Z+jets events and (e) gluon and quark induced diboson events. Multi-jet background is not shown as it is reduced to the percent level for the 1-lepton channel while its impact was found negligible for the 0- and 2-lepton channel.

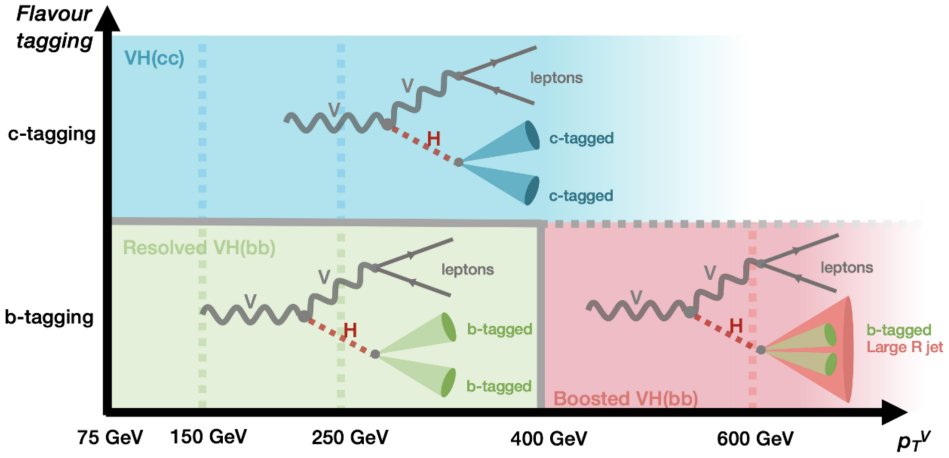


Figure 5.5: Analysis regimes considered in the $VH, H \to b\bar{b}/c\bar{c}$ analysis. The $VH, H \to b\bar{b}$ and $VH, H \to c\bar{c}$ regimes are separated through the requirement of b-and c-tags, while the resolved and boosted $VH, H \to b\bar{b}$ regimes are separated with a p_T^V cut of 400 GeV [62].

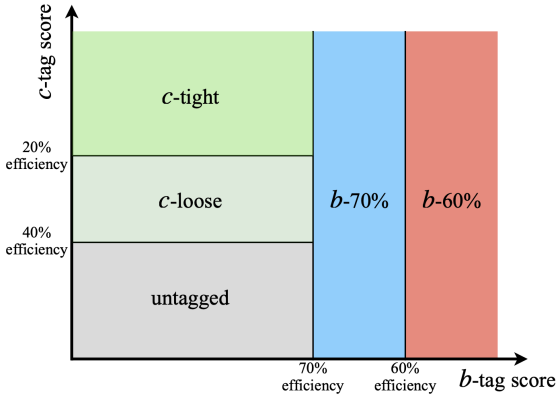


Figure 5.6: DL1r pseudo-continuous tagging algorithm working point definition [62].

channels. In addition, Standard Model di-boson processes, such as VZ with $Z \to b\bar{b}$ and $Z \to c\bar{c}$, are included as signal processes in dedicated cross-check analyses.

The dominant background contributions in both the resolved and boosted regimes arise from V+jets and $t\bar{t}$ production. The V+jets background is further decomposed into three components according to the jet-flavour content: $V + bb/cc$, $V + bc/bl/cl$, and V+light-flavour jets. The $t\bar{t}$ background is instead split into two categories, with the $t\bar{t} + bb$ component treated separately from the remaining dijet flavour final states.

The classification is performed at generator level according to the flavour of heavy hadrons associated with the reconstructed jets. Events containing two heavy-flavour jets (b or c) are assigned to the $V + bb/cc$ category, events with a single heavy-flavour jet to the $V + bc/bl/cl$ category, while events with no heavy-flavour jets are classified as V+light. In cases where multiple heavy flavours are present (e.g. bbc), the event is assigned to the category corresponding to the two heavy-flavour jets.

5.2.2 Flavour tagging

b- and c-jets are classified using a pseudo-continuous implementation of the DL1r tagging algorithm [63], which enables a coherent definition and calibration of both flavour-tagging cate-

gories. As illustrated in Figure 5.6, two working points are defined for both b- and c-tagging, resulting in a subdivision of the two-dimensional b- and c-score plane into five distinct bins to which each jet is assigned.

Two of these five bins correspond to the 60% and 70% b-tagging working points, with the 70% category defined excluding events contained in the 60% one. Similarly, two c-tagging bins are defined based on the charm-tagging discriminant: a *tight* working point selecting jets with high c-jet purity, and a *loose* working point with higher efficiency but lower purity, while the remaining bin collects untagged jets.

Events are selected using criteria that will be specified in Section 5.3.1, when the analysis presented in this thesis is introduced, since the two kinematic selection sets are identical and therefore deferred to that section. In $VH, H \to b\bar{b}$ analysis, signal $bb-$events are then selected in the resolved topology with the DL1r 70% efficiency working point, while the boosted topology uses the DL1r 85% working point.

5.2.3 Trigger Event Selection

In the $VH, H \to b\bar{b}/c\bar{c}$ analysis, a common trigger strategy is adopted for both final states to ensure consistency across channels [62]. The choice of trigger depends on the lepton multiplicity and the kinematic properties of the event.

In the 0-lepton channel, events are selected using the lowest unprescaled E_T^{miss} triggers. The online thresholds depend on the data-taking period: 70 GeV for 2015, 90 and 110 GeV for 2016, and 110 GeV for both 2017 and 2018. The increase in threshold in later years reflects the higher instantaneous luminosity and resulting trigger rates.

In the 1-lepton channel, events are selected using single-lepton triggers. Electron events are required to pass at least one of the lowest unprescaled single-electron triggers. For muon events, a two-regime strategy is adopted: for $p_T^V > 150$ GeV, E_T^{miss} triggers are used, as in the 0-lepton channel, in order to maximise the trigger efficiency in the boosted regime; for $p_T^V \leq 150$ GeV, the lowest unprescaled single-muon triggers are instead employed.

In the 2-lepton channel, a similar strategy is followed. Electron events rely on single-electron triggers, while in the muon channel the transition between single-muon and E_T^{miss} triggers is shifted to higher transverse momentum, occurring at $p_T^V = 250$ GeV. As shown in Ref. [64], the use of E_T^{miss} triggers in this high-p_T^V region leads to an increase in signal acceptance of about 5%.

For all lepton-based triggers, a trigger matching requirement is applied. This ensures that the reconstructed lepton in the event is geometrically and kinematically consistent with the object that fired the trigger. A detailed description of the trigger matching procedure is provided in Ref. [65].

5.2.4 Event Categorization

As summarized in Figure 5.7, events are further split into different categories, considering for all of them the 0-lepton, 1-lepton and 2-lepton channels separately.

Resolved $VH, H \to b\bar{b}$ regime signal region is populated by events with exactly two $b-$jets and is divided in different p_T^V intervals between 75 and 400 GeV, also separating events with 2 and 3 jets. 0-lepton and 2-lepton channels also consider the 4-jet category, requiring exactly 4 jets in the 0-lepton case and 4 or more in the 2-lepton one. An additional control region orthogonal to the signal one is defined for each case utilizing events with high ΔR between the Higgs candidate jets for determining the $V+$jets and $t\bar{t}$ events yields from data. In the 1-lepton channel the $W + bb$ contribution is controlled through a low-ΔR control region.

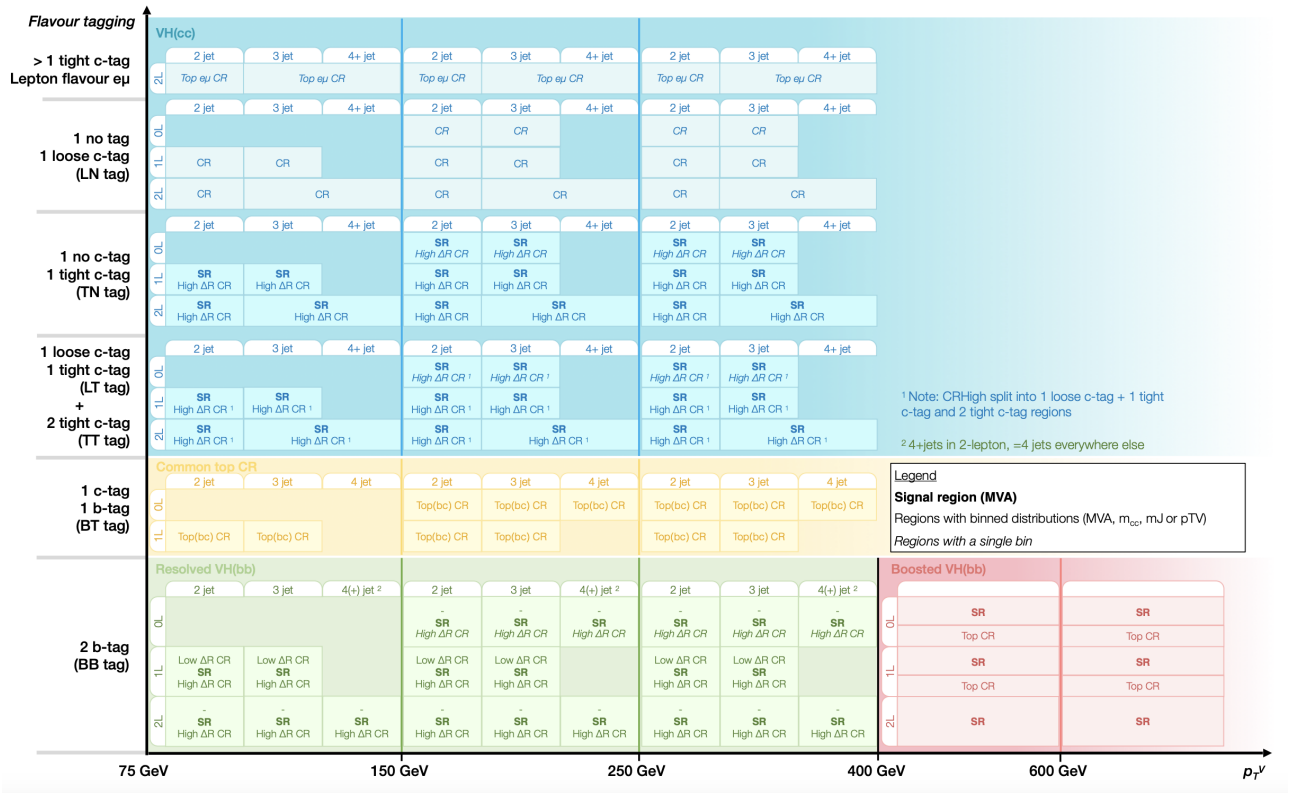


Figure 5.7: Analysis regions considered in the $VH(\to b\bar{b}, c\bar{c}$ analysis [62].

A similar categorization is defined for the $VH, H \to c\bar{c}$ analysis channel. Same p_T^V and jet multiplicity regions are defined, but keeping events with p_T^V >400 GeV in the baseline signal region as the overlap with the boosted $VH, H \to b\bar{b}$ region is small. Here, signal regions are required to have at least one tight $c-$tag jet, the different signal regions are defined using events which additionally have a second loose or tight $c-$tag jet. A second set of signal regions is defined by events with one correctly $c-$tag jet and untagged additional jets. As for the $VH(b\bar{b})$ channel, a corresponding high-ΔR control region is defined for each signal region.

Top control regions are shared between resolved $VH, H \to b\bar{b}$ and $VH, H \to c\bar{c}$ regimes. One $b-$tag and one tight $c-$tag jets are required in the 0-lepton and 1-lepton channels, while in the 2-lepton channels they are defined by requiring different lepton flavours (one electron and one muon).

The boosted $VH, H \to b\bar{b}$ categorizes events in two p_T^V regions, and signal regions require exactly two of three leading sub-jets of the leading large-R jet to be $b-$tagged. In the **0-lepton channel**, events are further categorized according to the number of additional small-R jets outside the large-R jet and the presence of reconstructed hadronic τ candidates. Four categories are produced: events with no additional small-R jets and no hadronic τ candidates; events with no additional small-R jets and one hadronic τ candidate; events with one additional small-R jet and no hadronic τ candidates; and events with one additional small-R jet and one hadronic τ candidate.

This separation reflects the different physical processes contributing to the selected sample. Additional small-R jets outside the large-R jet are often associated with heavy-flavour jets produced in top-quark decays and are therefore useful for identifying events entering the Top control region. The presence of hadronic τ candidates is relevant in the 0-lepton channel because events in which the W boson decays to a τ lepton, followed by a hadronic τ decay, can enter the selection without reconstructed electrons or muons. Separating these configurations allows a dedicated treatment of such topologies in the analysis.

In the **1-lepton channel**, the output is divided according to the presence of additional small-R jets outside the large-R jet. Events are therefore separated into two categories: events without additional small-R jets and events containing at least one additional small-R jet. This distinction is particularly relevant for identifying top-quark backgrounds, since the decay of $t\bar{t}$ pairs typically produces extra b-jets outside the Higgs candidate. Consequently, the category containing additional jets is used to populate the Top control region, while events without extra jets correspond to the signal region selection.

Finally, in the **2-lepton channel**, events are required to contain a dilepton pair compatible with the decay of a Z boson. In this case the background from top-quark production is already strongly suppressed by the dilepton selection, and therefore no dedicated control region is defined. The output consequently contains a single category corresponding to events with the Higgs candidate large-R jet and no additional small-R jets outside it.

5.2.5 Multi-Variate Analysis

Multivariate analysis (MVA) techniques are used to combine multiple input variables into a single discriminant to enhance the separation between signal and background. By exploiting correlations among observables, MVAs provide greater sensitivity than traditional cut-based selections, making them particularly effective in complex final states such as those involving heavy-flavour jets.

The MVA $VH, H \to b\bar{b}/c\bar{c}$ analyses is based on boosted decision trees (BDTs). A decision tree is a hierarchical classifier that splits the dataset recursively according to the values of input variables, creating a tree-like structure where each node represents a decision based on a single variable, and the leaves correspond to classification outcomes, such as signal or background. By combining multiple variables and exploiting their correlations, a single decision tree can capture complex patterns in the data. However, individual trees are prone to overfitting and can be sensitive to statistical fluctuations. The "boosted" part of BDTs [66] addresses this limitation by training an ensemble of trees sequentially, where each new tree focuses on events that were misclassified by the previous one. The outputs of all trees are then combined, usually through a weighted majority vote or a sum of tree scores, producing a robust discriminant with improved classification performance and reduced sensitivity to statistical noise.

BDTs are used both in the resolved and boosted regime of the $VH, H \to b\bar{b}$ and $VH, H \to c\bar{c}$ analyses, with a BDT trained for each region in which the events are categorized as described in the previous paragraph. The input variables are optimized for each leptonic channel in the boosted and resolved regime, as showed in Table 5.8.

Typically, it has been observed a 30%-50% improvement (mostly depending on the lepton channels) in sensitivity when using MVA algorithms compared to the mass-based analysis (where the sensitivity is computed inspecting directly the signal and background mass distributions as done in the preliminary studies to the $\mu^{post-tag}$ measurement in the GN2X calibration), explained by the fact that an MVA algorithm uses several relevant input variables, exploiting correlations between them and learning characteristics specific to different regions of the phase space which are more complex than a simple mass-based analysis.

The development of each BDT follows a standard machine learning procedure: the algorithm is first trained to distinguish signal from background events using MC simulations where the true event category is known, then its performance are evaluated on a statistically independent sample in terms of ROC curve. BDT's hyper-parameters, such as the number of trees, the maximum depth of each of them, the learning rate and the number of cuts to be applied on each variable are optimized for each leptonic channel and regime for both the final states, as listed in Tables 5.9 and 5.10.

Specifically, the hyper-parameters optimized for each configuration in order to maximize

		$VH(\to b\bar{b}, c\bar{c})$ Resolved			$VH(\to b\bar{b})$ Boosted		
Variable	Name	0-lepton	1-lepton	2-lepton	0-lepton	1-lepton	2-lepton
$m_{j_1 j_2}$ or m_J	mBB / mJ	×	×	×	×	×	×
$m_{j_1 j_2 j_3}$	mBBJ	×	×	×			
$p_{\mathrm{T}}^{j_1}$ or $p_{\mathrm{T}}^{j_1,\mathrm{trk}}$	pTB1 / pTBTrkJ1	×	×	×	×	×	×
$p_{\mathrm{T}}^{j_2}$ or $p_{\mathrm{T}}^{j_2,\mathrm{trk}}$	pTB2 / pTBTrkJ2	×	×	×	×	×	×
$p_{\mathrm{T}}^{\mathrm{j_3,trk}}$	pTBTrkJ3				×	×	×
$\sum_{i\neq 1,2} p_{\mathrm{T}}^{j_i}$	sumPtAddJets	×	×	×			
$\Delta R(\vec{j}_1, \vec{j}_2)$ or $\Delta R(j_{1,\mathbf{trk}}, j_{2,\mathbf{trk}})$	dRBB / deltaRbTrkJbTrkJ	×	×	×	×	×	×
$\lvert\Delta\eta(\vec{j}_1, \vec{j}_2)\rvert$	dEtaBB	×					
$\mathrm{bin}_{\mathrm{DL1r}}(j_1)$	bin_btagB1 / bin_bTagBTrkJ1	×	×	×	×	×	×
$\mathrm{bin}_{\mathrm{DL1r}}(j_2)$	bin_btagB2 / bin_bTagBTrkJ2	×	×	×	×	×	×
p_{T}^{V}	pTV	$\equiv E_{\mathrm{T}}^{\mathrm{miss}}$	×	×	$\equiv E_{\mathrm{T}}^{\mathrm{miss}}$	×	×
$E_{\mathrm{T}}^{\mathrm{miss}}$	MET	×	×		×	×	
$E_{\mathrm{T}}^{\mathrm{miss}}/\sqrt{S_{\mathrm{T}}}$	METSig			×			
$\lvert\Delta y(\vec{V}, \vec{H}_{\mathrm{cand}})\rvert$	dYVBB/deltaYVJ		×	×		×	×
$\lvert\Delta\phi(\vec{V}, \vec{H}_{\mathrm{cand}})\rvert$	dPhiVBB/absdeltaPhiVJ	×	×	×	×	×	×
$\min[\Delta\phi(\vec{\ell}, \vec{j}_1 \text{ or } \vec{j}_2)]$	dPhiLBmin		×				
m_{eff}	MEff	×					
m_T^W	mTW		×				
m_{top}	Mtop		×				
$m_{\ell\ell}$	mLL			×			
$\cos\theta(\vec{\ell}^-, \vec{Z})$	cosThetaLep			×			×
$(p_{\mathrm{T}}^{\ell_1} - E_{\mathrm{T}}^{\mathrm{miss}})/p_T^W$	lepPtBalance					×	
p_{T}^{ℓ}	pTL					×	
N(track-jets in J)	NMatchedTrackJetLeadFatJet				×	×	×
N(add. small R-jets)	NAdditionalCaloJets				×	×	×
Colour	Colour				×	×	×
$\min\{\Delta R(b, j)\}$	minDRBjets	×	×				

Figure 5.8: MVA variables used for each leptonic channel in the resolved and boosted regimes for the $VH(b\bar{b})$ and $VH(c\bar{c})$ analyses [62].

	Resolved $VH, H \to b\bar{b}$			Boosted $VH, H \to b\bar{b}$		
Settings	0-lepton	1-lepton	2-lepton	0-lepton	1-lepton	2-lepton
Boost type	Gradient boost	Gradient boost	Gradient boost	Adaboost	Adaboost	Adaboost
Number of trees	200	600	200	800	800	400
Maximum depth	3	4	4	3	3	3
Learning rate (β)	0.5	0.5	0.5	0.35	0.35	0.3
Number of cuts	100	100	100	60	60	100
Minimum node size	5%	5%	5%	2%	2%	7%
Separation method	Gini index	Gini index	Gini index	Gini index	Gini index	Gini index
Pruning method	No pruning	No pruning	No pruning	No pruning	No pruning	No pruning

Figure 5.9: BDT hyper-parameters used for each leptonic channel for $VH, H \to b\bar{b}$ analysis for resolved and boosted regime [62].

	$VH, H \to c\bar{c}$		$VH, H \to c\bar{c}$, Diboson as signal
Settings	0-, 1- and most 2-lepton regions	2-leton, ≥ 3-jet, low pTV	0-, 1- and 2-lepton
Boost type	Gradient boost	Adaboost	Adaboost
Number of trees	600	200	200
Maximum depth	4	4	4
Learning rate (β)	0.5	0.15	0.15
Number of cuts	100	100	100
Minimum node size	5%	5%	5%
Separation method	Gini index	Gini index	Gini index
Pruning method	No pruning	No pruning	No pruning

Figure 5.10: BDT hyper-parameters used for each leptonic channel for $VH, H \to c\bar{c}$ analysis.

the discrimination performances are defined as follows:

- The number of trees represents how the model is deep and is defined in order to find a balance between overtraining and performance of the model;
- The learning rate controls the contribution of each successive tree to the final model, effectively slowing down the learning to improve stability and reduce overfitting;
- The maximum depth of the cell tree represent the successive cuts that each decision tree is allowed to perform in order to further split the phase space into signal and background dominated sub-regions;
- The number of cuts, or number of grid points, is used for finding the optimal cut of a variable for a node splitting, the finer the granularity is the better should be the cut at the expense of computation time and overtraining;
- The minimum node size is the minimum percentage of training events required in a leaf node. It avoids having parts of the phase space that are too finely split by the BDT which can cause overtraining due to low statistic;
- The separation method represent the criteria used for dividing signal from background events. The used *Gini index* method is defined as

$$\text{Gini} = p(1-p) \tag{5.1}$$

with p purity of a node defined as $p = S/(S+B)$. S and B represent respectively the weighted sum of signal and background events contained in a given node during the training and lead to a maximum value of the Gini index of 0.5 when the amount of signal and background events in the node is the same. For each node, the phase space is divided in 2 child nodes by a cut, the maximum separation between the parent node and the two child nodes is then defined as:

$$\Delta\text{Gini} = -\text{Gini}_{\text{child}_1} - \text{Gini}_{\text{child}_2}. \tag{5.2}$$

The cut corresponding to the highest ΔGini separation is applied, and the iteration continues for the child nodes until the number of maximum cut is reached, then, depending on the purity of the node, all events reachinf the node will all be either classified as signal or background ones;

- Pruning is a technique used in decision trees to remove branches that provide little or no improvement in classification, reducing the tree's complexity and helping to prevent overfitting. Here, since the maximal depth for the BDT training remains small in all the configurations, there is no reason to apply the pruning method.

BDT performance are evaluated using the Asimov significance, which is computed from the BDT output distributions as follows:

$$S = \sqrt{\sum_i 2\left[(s_i + b_i)\ln(1 + \frac{s_i}{b_i}) - s_i\right]}, \tag{5.3}$$

with s_i and b_i being the total number of signal and background events contained in the $i-$bin of the BDT output distribution.

As raw output of a BDT is typically finely binned (e.g., 500 bins), suboptimal large statistical fluctuations are associated to many bins. In particular, bins in the signal-rich region may have limited statistical power, while those in background-dominated regions contribute little to discrimination. To address this, a variable-binning procedure, known as *Transformation D*, is applied to the BDT output before calculating the significance.

The purpose of this procedure is to merge the original bins into a smaller set, optimizing the balance between statistical precision and signal-background separation. The algorithm begins at the most signal-like bin (highest BDT score) and iteratively merges adjacent bins until a predefined criterion on the statistical uncertainty of the merged bin is satisfied. This results in a distribution with finer binning in the high-score region, where the signal is concentrated and the shape of the BDT output is most informative.

A general formula describing such merging, also referred to as remapping, is given by:

$$Z_i(j,k) = Z(z_s, z_b, N_s, N_b, n_s(j,k), n_b(j,k)), \tag{5.4}$$

with $Z_i(j,k)$ representing the $i-$bin of the transformed distribution obtained merging all bins from the original distribution between the $j-$ and the $k-$bins included. z_s and z_b are the parameters used to tune the merging, controlling the number of bins in the final transformed distribution that are allocated to signal and background events respectively.N_s and N_b quantify the total number of signal and background events, while, $n_s(j,k)$ and $n_b(j,k)$ represent the total number of signal and background events contained in the interval between the $j-$ and $k-$bins of the original distribution.

The general remapping formula described above, in the case of the transformation D becomes:

$$Z_i(j, k_{\text{last}}) = z_s \frac{n_s(j, k_{\text{last}})}{N_s} + z_b \frac{n_b(j, k_{\text{last}})}{N_b}, \tag{5.5}$$

with k_{last} last bin on the right of the original distribution from which the transformation starts merging one after the other the bins on its left. Bins are merged until the statistical uncertainty to each of them related is smaller than 20%. The distribution after transformation D results to have variable size bins, but for aesthetic reasons are generally plotted all with the same size.

5.2.6 Systematic Uncertainties

The precision of the $VH, H \to b\bar{b}/c\bar{c}$ measurement is limited by a variety of systematic uncertainties, which are incorporated into the final likelihood fit as constrained nuisance parameters. These uncertainties can be broadly classified into two categories: experimental, which arise

from the performance and calibration of the detector and reconstructed objects, and theoretical, which are related to the modelling of signal and background processes in Monte Carlo simulations.

Among the experimental uncertainties, the dominant contributions stem from jet calibration and flavour-tagging performance. For jets, this includes uncertainties on the Jet Energy Scale (JES) and Jet Energy Resolution (JER), as well as the corresponding mass scales and resolutions for both small-R and large-R jets. Flavour-tagging uncertainties, which account for the correction of b-tagging, c-tagging, and light-jet mistagging efficiencies in simulation to match data, represent a particularly significant source of uncertainty in this analysis.

Additional experimental uncertainties arise from several sources. These include the reconstruction and identification efficiencies of leptons, their momentum scale and resolution, and the measurement of missing transverse momentum (E_T^{miss}), which depends on both the calibration of reconstructed objects and the modelling of soft contributions. The effects of pileup, which are corrected via reweighting procedures in simulation, also introduce systematic uncertainties. Finally, the measurement of the integrated luminosity carries an uncertainty that affects all processes normalized to the total delivered dataset.

Together, these experimental uncertainties are carefully evaluated and propagated through the analysis, as they play a crucial role in defining the overall precision of the Higgs boson decay measurements in both the $b\bar{b}$ and $c\bar{c}$ channels.

5.2.7 Theoretical Uncertainties

A major source of systematic uncertainty comes from the theoretical modelling of both signal and background processes in the Monte Carlo (MC) simulations. These uncertainties are evaluated by applying a set of prescribed variations to the nominal event generators. The main modelling uncertainties are obtained by comparing the nominal MC samples with alternative samples that employ different matrix-element generators (e.g., MadGraph5_aMC@NLO) or different parton shower and hadronisation models (e.g., Herwig7).

Additional uncertainties are included to account for missing higher-order corrections, which are estimated by varying the renormalisation (μ_R) and factorisation (μ_F) scales, as well as the effects of initial- and final-state radiation (ISR/FSR) by adjusting the corresponding parameters in the parton shower simulation. For the V+jets background, further uncertainties are applied to the relative normalisation of the different heavy-flavour components to account for potential mismodelling of the flavour composition.

These theoretical uncertainties are propagated through the analysis and incorporated as constrained nuisance parameters in the final likelihood fit, ensuring a robust estimation of their impact on the measurement of the $VH, H \to b\bar{b}/c\bar{c}$ signal.

A summary of all the uncertainties included into the analysis, together with their contribution, is reported in Figure 5.11.

5.2.8 Fit Formalism

The signal and background yields are determined, after selecting the events, splitting them in regions and for each leptonic channel and region training a BDT for discriminating signal from background events, through maximum likelihood fits to the obtained BDT output distributions for each topology, channel and signal region presented in previous sections.

In general, having a distribution of interest of a variable x with N bins containing respectively $\mathbf{n} = (n_1, .., n_N)$ events to be fitted, then the expectation value $E[n_i]$ for the content n_i of the i^{th} bin, in the context of the signal-plus-background hypotesis, can be expressed as

$$E[n_i] = \mu s_i + b_i, \tag{5.6}$$

Source of uncertainty		σ_μ			
		$VH, H \to b\bar{b}$	$WH, H \to b\bar{b}$	$ZH, H \to b\bar{b}$	$VH, H \to c\bar{c}$
Total		0.153	0.204	0.216	5.31
Statistical		0.097	0.139	0.153	3.94
Systematic		0.118	0.149	0.153	3.57
Statistical uncertainties					
Data statistical		0.090	0.129	0.139	3.67
$t\bar{t}$ $e\mu$ control region		0.009	0.014	0.027	0.08
Background floating normalisations		0.034	0.049	0.042	1.24
Other VH floating normalisation		0.007	0.018	0.014	0.33
Simulation samples size		0.023	0.033	0.030	1.62
Experimental uncertainties					
Jets		0.027	0.035	0.030	1.02
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.010	0.005	0.021	0.23
Leptons		0.003	0.002	0.010	0.25
b-tagging	b-jets	0.020	0.018	0.026	0.29
	c-jets	0.013	0.017	0.012	0.73
	light-flavour jets	0.005	0.008	0.008	0.66
Pile-up		0.008	0.017	0.002	0.23
Luminosity		0.006	0.007	0.006	0.08
Theoretical and modelling uncertainties					
Signal		0.076	0.074	0.101	0.72
Z + jets		0.042	0.018	0.081	1.77
W + jets		0.054	0.087	0.026	1.42
$t\bar{t}$ and Wt		0.018	0.033	0.018	1.02
Single top-quark (s-, t-ch.)		0.010	0.018	0.002	0.16
Diboson		0.033	0.039	0.049	0.52
Multijet		0.005	0.010	0.005	0.55

Figure 5.11: Breakdown of contributions to the uncertainties in the fitted signal strength values for each category [62].

with s_i and b_i respectively the number of signal and background events expected for the i^{th} bin and μ signal strength. s_i and b_i are computed from the respective probability density functions (PDF) f_s and f_b defined as

$$s_i = s_{\text{tot}} \int_{\text{bin } i} f_s(x, \theta_s) dx; \qquad b_i = b_{\text{tot}} \int_{\text{bin } i} f_b(x, \theta_b) dx, \tag{5.7}$$

where s_{tot} and b_{tot} represent the total number of signal and background events as predicted by the simulation and θ_s and θ_b are the set of parameters controlling the shapes of the PDFs. As explained in the Chapter 4, the parameter μ is the so-called signal strength, defined as the ratio between the observed and expected number of signal events.

$$\mu := \frac{\sigma \cdot BR}{\sigma_{\text{SM}} \cdot BR_{\text{SM}}} = \frac{N_{sig}^{obs}}{N_{sig}^{exp}}. \tag{5.8}$$

While fitting, the only parameter of interest (POI) is the signal strength, while $\Theta = (\theta_s, \theta_b, \theta_{\text{tot}})$ defined as set of nuisance parameters (NP), i.e. the set of parameters which must be accounted for in the hypotesis testing but which are not of interest.

In order to measure the signal strenght μ in the $VH, H \to b\bar{b}/c\bar{c}$ analysis, a binned likelihood function is used on the BDT output distributions. The likelihood function can be decomposed in the product of three different likelihoods

$$\mathcal{L}(\mu, \theta) = \mathcal{L}_{\text{EML}}(\mu, \alpha, \gamma, \tau) \cdot \mathcal{L}_{\text{syst}}(\alpha) \cdot \mathcal{L}_{\text{stat}}(\gamma) \tag{5.9}$$

with:

- μ the parameter of interest,
- $\theta = (\alpha, \gamma, \tau)$ the nuisance parameters respectively taking into account the experimental and modelling uncertainties (α), the MC statistical uncertainty on the sum of background processes in each bin (γ) and the floating normalizations for the different background processes (τ). Conventionally $\alpha_j = 0$ is for nominal prediction while $\alpha_j = \pm 1$ are associated to the $\pm 1\sigma$ variations related to a given uncertainty which can affect one or several processes. γ_i parameters are used for multiplying the sum of the backgrounds b_i in each bin in order to allow the $\gamma_i b_i$ quantity to vary within the background MC statistical uncertainty of the i^{th} bin. τ parameters are free parameters determined during the fit without any prior.
- $\mathcal{L}_{\text{EML}}$ is the extended maximum likelihood, defined as the product of the Poisson distributions in each bin of signal-plus-background events as they are assumed to occur independently and the average rate at which they occur is independent of any occurrences

$$\mathcal{L}_{\text{EML}}(\mu, \alpha, \gamma, \tau) = \prod_{i=1}^{N_{\text{bins}}} \frac{(\mu s_i(\alpha) + b_i(\alpha, \gamma, \tau))^{n_i}}{n_i!} \cdot e^{-(\mu s_i(\alpha) + b_i(\alpha, \gamma, \tau))}. \tag{5.10}$$

- $\mathcal{L}_{\text{syst}}$ is the likelihood associated to the experimental and modelling uncertainties, computed, under the assumption of having not correlated systematic uncertainties, as the product of Gaussian distributions

$$\mathcal{L}_{\text{syst}}(\alpha) = \prod_{j=1}^{N_{\text{syst}}} \frac{e^{-\alpha_j^2/2}}{\sqrt{2\pi}}. \tag{5.11}$$

- $\mathcal{L}_{\text{stat}}$ is the likelihood associated to the MC statistical uncertainty on the sum of all backgrounds and is obtained from the product of the (3-dimensional) Γ function over all bins, representing the probability for each bin i to have b_i background events predicted when the true number of events is $\gamma_i b_i$ [1]

$$\mathcal{L}_{\text{stat}}(\gamma) = \prod_{i=1}^{N_{\text{bins}}} \Gamma(\gamma_i b_i, b_i + 1, 1) = \prod_{i=1}^{N_{\text{bins}}} \frac{(\gamma_i b_i)^{b_i}}{\Gamma(b_i + 1)} \cdot e^{-\gamma_i b_i}, \tag{5.12}$$

with

$$\Gamma(\mu, \alpha, \beta) = \frac{\beta^\alpha \mu^{\alpha-1}}{\Gamma(\alpha)} \cdot e^{-\beta\mu}, \tag{5.13}$$

and the so called (1-dimensional) Γ function defined as $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$.

The best agreement between data and simulation is found when maximizing the likelihood function. Two different types of fits can be performed: a conditional fit, when the likelihood is maximized for a particular fixed value of the signal strength and only the NPs are allowed to vary, and the unconditional fit, in which no parameters are fixed while maximizing the likelihood, allowing to measure the signal strength in data.

5.2.9 Final Run-2 $VH, H \to b\bar{b}/c\bar{c}$ analysis results

The MVA output distributions are used to extract the signal and background yields through a simultaneous maximum likelihood fit performed across all topologies, channels, and signal regions. The fit exploits the full shape information of the BDT discriminant and incorporates the effects of systematic uncertainties via nuisance parameters constrained by their corresponding prior distributions. In this framework, the signal strengths are treated as free parameters of interest, while background normalizations and shape variations are profiled in the fit. The final measurements are obtained from the best-fit values of the parameters of interest and their associated uncertainties, derived from the profile likelihood as explained in detail in Section 5.2.8.

The use of BDTs for signal-to-background discrimination significantly enhances the sensitivity of the analysis and was a key factor in enabling the analysis to achieve the first observation of the individual WH production process with subsequent Higgs boson decays to b-quarks.

In the WH channel, the measurement resulted in an observed (expected) significance of 5.3 (5.5) standard deviations. The ZH, $H \to b\bar{b}$ process, treated independently, also reached a high level of sensitivity, with an observed (expected) significance of 4.9 (5.6) standard deviations.

Thanks to the orthogonal division of phase space to distinguish $H \to b\bar{b}$ and $H \to c\bar{c}$ regions, the analysis results in an observed (expected) upper limit on the $H \to c\bar{c}$ signal strength of 11.5 (10.6) times the SM prediction at 95% confidence level (CL) [62].

With the full Run-2 dataset, combining simultaneously the two channels, the fit result determined the following signal strengths:

$$\mu_{VH}^{b\bar{b}} = 0.92^{+0.16}_{-0.15} = 0.92^{+0.10}_{-0.10}(\text{stat.})^{+0.13}_{-0.11}\ (\text{syst.}), \tag{5.14}$$

$$\mu_{VH}^{c\bar{c}} = 1.0^{+5.4}_{-5.2} = 1.0^{+4.0}_{-3.9}(\text{stat.})^{+3.7}_{-3.5}\ (\text{syst.}). \tag{5.15}$$

Results are reported in Figure 5.12 both for $VH, H \to b\bar{b}$ and $VH, H \to c\bar{c}$.

Beyond the inclusive measurements, the maximum likelihood fit is also used to extract differential cross sections for both WH and ZH production within the Simplified Template

[1] Those terms are in principle Poisson terms, as $\Gamma(n+1) = n!$ for any positive integer n, but since b_i is not necessarily an integer, the Γ function is needed in the denominator instead of the factorial term.

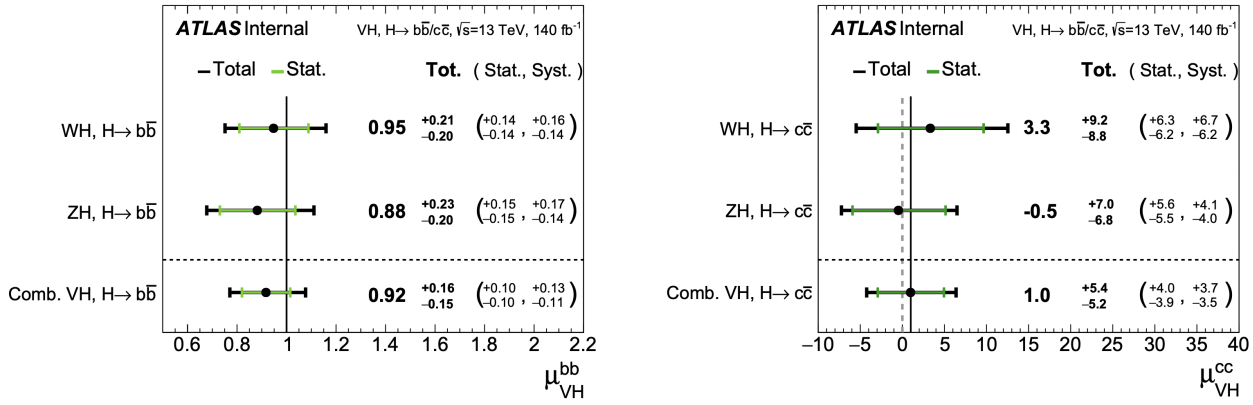


Figure 5.12: Fitted signal strength and uncertainties from the lepton channel decorrelated and combined fit, left is $VH, H \to b\bar{b}$ and right is $VH, H \to c\bar{c}$.

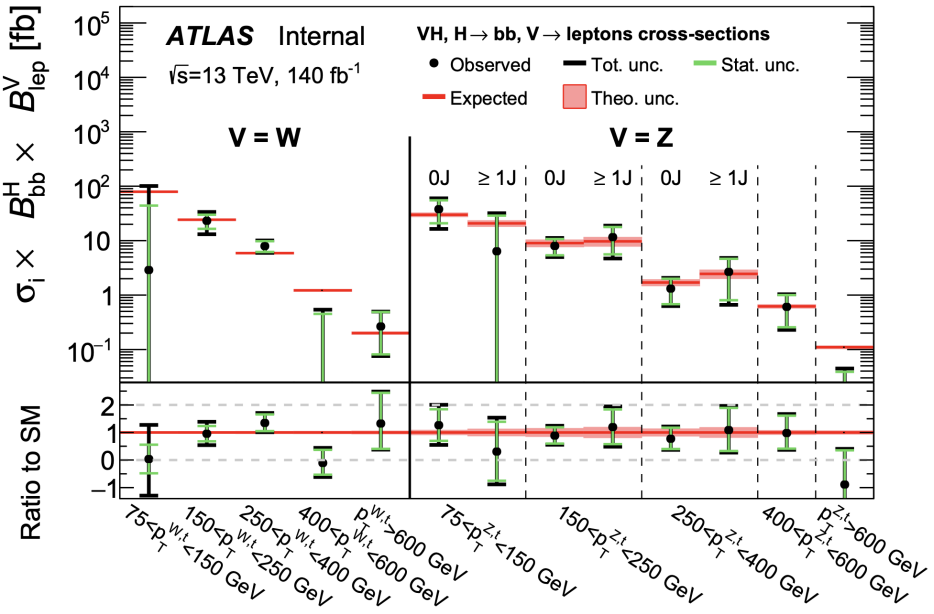


Figure 5.13: Measured $VH, V \to$ leptons cross section times $H \to b\bar{b}$ BR's in the STXS extended scheme [62].

Cross Section (STXS) framework [67]. These measurements are performed in 13 distinct fiducial regions and further subdivided according to the number of additional jets in the final state.

Figure 5.13 presents the results obtained using this extended scheme, shown in terms of the production cross section times the branching ratio for each fiducial region. The measured values are found to be in good agreement with the Standard Model predictions.

In addition to the CL result for the rare $H \to c\bar{c}$ decays, a direct constraint on the charm Yukawa coupling modifier, κ_c, was interpreted through the so called $\kappa-$framework. Assuming all other Higgs couplings are at their SM values, the analysis sets an observed upper limit of $|\kappa_c|$ <4.2 at 95% CL. A combination in a global fit of both $H \to b\bar{b}$ measurement and $H \to c\bar{c}$ search result in a constraint placed on the absolute value of the ratio of the coupling modifiers of $|\kappa_c/\kappa_b| < 3.6$ at 95% CL, providing an experimental confirmation of the fact that the Higgs boson couples more weakly to charm quarks than to bottom ones (as the predicted mass-proportionality of the couplings in the SM).

5.3 Implementation of the Calibrated GN2X Tagger in the Boosted $VH, H \to b\bar{b}$ Analysis

Building upon the framework established in the ATLAS Run-2 $VH, H \to b\bar{b}/c\bar{c}$ analysis, the following sections present the implementation of the GN2X boosted $b\bar{b}$ tagger within the $VH, H \to b\bar{b}$ analysis. The goal of this work is to assess the impact of the GN2X tagger on the sensitivity of the boosted analysis and to quantify its improvement over previous tagging strategies.

While the overall event selection largely follows the same kinematic criteria as in the ATLAS Run-2 $VH, H \to b\bar{b}/c\bar{c}$ analysis, the key distinction lies in the strategy used to identify Higgs candidates: instead of relying on the DL1r pseudo-continuous tagger, the GN2X tagger, along with its corresponding scale factors (SFs), is now applied directly to large-R jets at its 70% working point, providing a more targeted and efficient identification of signal $b\bar{b}$ decays in the boosted regime.

The analysis workflow mirrors the last analysis framework also in its structure: after the event selection and the analysis region definition, signal events are used in a preliminary multivariate analysis to produce BDT output scores that then have to be used within the statistical fit.

In the following sections, the construction of the input samples, the integration of the GN2X tagger and its SFs, the MVA setup, and the propagation of its output to the statistical analysis are described in detail. Finally, the improvements in signal significance achieved through the inclusion of the GN2X tagger are quantified, demonstrating its impact on the sensitivity of the boosted $VH, H \to b\bar{b}$ analysis.

5.3.1 Boosted events selection

The analysis follows the same overall channel division as the previous $VH, H \to b\bar{b}$ study presented in Section 5.2.4, separating events into 0-, 1-, and 2-lepton categories, each targeting a specific decay mode of the associated vector boson. Objects such as leptons, small-R jets, and missing transverse energy are reconstructed following the standard procedures established in the older analysis. Same trigger strategy, as explained in Section 5.2.3, is also used.

As this analysis focuses only on the boosted regime, the Higgs boson decay products results to be collimated, and the $b\bar{b}$ final state is reconstructed as a single large-R jet with radius parameter $R = 1.0$, which serves as the Higgs candidate.

The event selection procedure begins with a common set of baseline criteria aligned with the previous analysis ones applied to all leptonic channels to ensure a consistent and high-quality starting sample. These baseline selections require the presence of at least one large-R fat jet with transverse momentum $p_T > 200$ GeV and pseudorapidity $|\eta| < 2.0$, serving as the boosted Higgs candidate, and missing transverse energy (MET) greater than 200 GeV.

To suppress QCD background, the azimuthal separation between the MET vector and the leading small-R jets must satisfy $\Delta\phi(\text{MET}, j) > 30$ř for all central jets with $p_T > 70$ GeV.

As in the previous analysis, additional channel-specific selections are then applied to each channel:

- **0-lepton channel:** Events must contain no isolated electrons or muons, corresponding to the fully invisible decay of the associated vector boson $Z \to \nu\nu$;
- **1-lepton channel:** Exactly one isolated electron or muon with $p_T > 25$ GeV is required, consisted with the semileptonic vector boson decays;
- **2-lepton channel:** Exactly two isolated, opposite sign, electrons or muons, each with $p_T > 25$ GeV, corresponding to the decay $Z \to \ell^+\ell^-$.

A special case arises in the 0-lepton channel when the W boson produced in WH events decays into a tau lepton and a neutrino ($W \to \tau\nu$), and the tau subsequently decays hadronically. In such events no electron or muon is reconstructed in the final state, and so the event naturally satisfies the 0-lepton selection. However, hadronically decaying taus can be misidentified as jets, which are also expected from the Higgs boson decay products. As a consequence, these events can enter the 0-lepton event sample and minic the signal topology.

To account for this effect, a hadronic tau split is applied to events in the 0-lepton channel. Events containing at leas one reconstructed hadronic tau candidate are treated separately, imposing on them an additional selection requiring the transverse mass of the reconstructed W boson, defined using the tau candidate and the missing transverse energy as $m_T^W = \sqrt{2p_T^\tau E_T^{\text{miss}}(1 - cos\Delta\phi(\tau, E_T^{\text{miss}}))}$, to satisfy

$$m_T^W > 10 \text{ GeV} \tag{5.16}$$

in order to suppress the contamination from fake tau candidates while retaining events consistent with genuine $W \to \tau\nu$ decays.

In addition, for events containing a reconstructed hadronic tau, the transverse momentum of the vector boson is reconstructed differently from the standard 0-lepton case. Instead of being represented solely by the missing transverse energy, the vector boson transverse momentum is defined as the vectorial sum of the missing transverse momentum and the transverse momentum of the leading hadronic tau candidate:

$$\vec{p}_T^{\,V} = \vec{E}_T^{\text{miss}} + \vec{p}_T^{\,V}. \tag{5.17}$$

Therefore, in events with reconstructed hadronic taus, the quantity $\vec{p}_T^{\,V}$ is not directly equivalent to the missing transverse energy, but instead includes the visible momentum carried by the hadronic tau decay products. This treatment ensures a more accurate reconstruction of the underlying W boson kinematics and improves the modelling of these events within the 0-lepton analysis.

5.3.2 Boosted Higgs Reconstruction

After passing the event-level selections, the analysis proceeds to the boosted topology reconstruction. The leading large-R jet is identified as the Higgs candidate and analyzed together with its associated track jets.

The identification of the Higgs candidate is performed using the GN2X tagger, which operates directly on the reconstructed large-R jet. In principle, this tagging approach does not require the explicit use of track jets, since the GN2X algorithm directly evaluates the large-R jet properties to assign a tagging score. However, in this preliminary implementation of the tagger in the analysis, track jets associated with the large-R jet are used for a specific purpose related to the calibration of the tagging efficiency. As discussed in Chapter 4, the GN2X calibration available for this study was derived only for signal events. Therefore, the corresponding scale factors must be applied only to simulated events that are compatible with the expected $H \to b\bar{b}$ decay topology. To identify such events, track jets matched to the large-R jet are examined, and the scale factors are applied only to signal $VH, H \to b\bar{b}$ and di-boson events containing two truth $b-$jets inside the large-R one.

The GN2X 70% working point is then applied for selecting signal events. For events passing this requirement, the event weight is corrected by applying the appropriate scale factor for the large-R jet, which depends on the transverse momentum of the Higgs candidate and is taken from Table 4.15. This correction is applied only to the subset of signal and diboson events satisfying the two-b-jet requirement described above. In addition, in 0- and 1-lepton

channels, for events with an additional $b-$tagged small-R jet, it will contribute to the final event reweighting through their corresponding b-tagging GN2 scale factors. The final scale factor applied to the selected tagged event is therefore defined as the product of the correction associated with the large-R jet and the scale factors of the small-R jets outside the Higgs candidate:

$$SF_{\text{event}} = SF_{\text{large-R}}(p_T^H) \times \prod_{j \in \text{small-R jets outside}} SF_j. \tag{5.18}$$

This procedure ensures that the tagging efficiency of the boosted Higgs candidate and the flavour tagging of additional jets in the event are consistently corrected when comparing simulation with data.

In addition to the track jets associated with the Higgs candidate, small-R calorimeter jets that are not overlapping with the large-R jet are also considered in the event. These jets serve two main purposes. First, they help identify additional heavy-flavour activity, which is characteristic of background processes such as top-quark production. Second, they are used to define analysis regions, separating events into signal and control regions according to the number of additional jets and their flavour tagging properties. This categorization improves the ability to constrain background contributions in the statistical analysis.

5.3.3 Boosted events categorization

In the boosted analysis, selected events are categorized according to their lepton multiplicity (0L, 1L, 2L) and further separated into signal and control regions to optimize sensitivity and constrain the main backgrounds in the statistical analysis.

A key difference with respect to the previous boosted analysis concerns the transverse momentum of the reconstructed vector boson, p_T^V. While the previous implementation restricted the boosted selection to $p_T^V > 400$ GeV, in this work the threshold is lowered to $p_T^V > 250$ GeV. This extension is motivated by the improved performance of the GN2X tagger, which provides sufficient discrimination between signal and background even in this intermediate transverse momentum regime, allowing additional signal events to be included and improving the overall sensitivity of the analysis.

Within each leptonic channel, events are classified based on the presence of the Higgs boson candidate and additional small-R jets outside the large-R one.

Signal regions (SRs) are defined for each leptonic channel requiring an Higgs candidate passing the GN2X 70% $bb-$tagging requirement, with no additional tagged jets outside the large-R Higgs candidate.

For the 0-lepton and 1-lepton channels, a dedicated top control region (TopCR) is defined to constrain backgrounds from top-quark production. Events entering the TopCR are required to contain at least one b-tagged small-R jet outside the large-R Higgs candidate. This requirement exploits the typical topology of $t\bar{t}$ events, where the decay of the two top quarks produces additional b-quarks that may appear as separate jets in the event. As a result, $t\bar{t}$ processes, which represent one of the dominant backgrounds in these channels, are significantly enriched in this region.

Figure 5.14 illustrates an example of such a topology, where part of the hadronically decaying top quark is reconstructed inside a large-R jet, while an additional b-tagged track jet is produced in the decay of the second top quark. The TopCR therefore provides a selection that is orthogonal to the signal region while enhancing the contribution of top-quark events, allowing their normalization to be constrained in the statistical fit.

In the 2-lepton channel, no dedicated control region is defined, since the requirement of a dilepton system compatible with a Z boson already strongly suppresses the contribution from

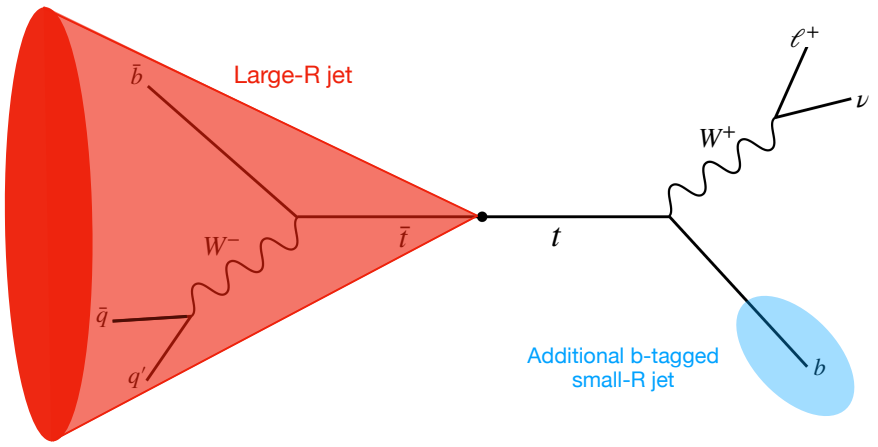


Figure 5.14: $t\bar{t}$ event in which the decay products of one top quark, with a hadronic-decay W, are reconstructed in a large-R jet (red) and an additional b-tagged track-jet is produced in the other top-quark decay (light blue).

top-quark processes.

All events written to the output satisfy the boosted Higgs reconstruction requirements, namely the presence of a large-R jet identified as the Higgs boson candidate and containing two b-tagged track jets. These events therefore correspond to configurations compatible with the $H \to b\bar{b}$ decay in the boosted regime. For each leptonic channel events are further categorized according to the same criteria presented in Section 5.2.4 for what concern the boosted regime of the analysis.

This output structure provides a natural mapping to the boosted signal and control regions defined in the analysis, allowing events with additional heavy-flavour activity or hadronic τ candidates to be treated separately from the cleaner signal-like configurations.

Table 5.1: Structure of the boosted CxAODReader outputs for each leptonic channel. All events contain a large-R Higgs candidate with two b-tagged track jets. Categories are defined according to the presence of additional small-R jets and hadronic τ candidates.

Channel	**Additional small-R jets**	**Hadronic τ**	**Typical usage**
0L	0	0	Signal-like region
0L	0	1	Events with hadronic τ from $W \to \tau\nu$
0L	1	0	Top-enriched topology (TopCR candidate)
0L	1	1	Top-like topology with hadronic τ
1L	0	–	Signal region
1L	≥ 1	–	Top control region
2L	0	–	Signal region

Table 5.2: `CxAODReader` output structure, each leptonic channel is processed individually and further categorized in different final states.

Table 5.1 summarizes the structure of the boosted outputs produced by the `CxAODReader`. This categorization reflects the different physical topologies present in the selected events and provides a direct mapping to the signal and control regions used in the statistical analysis.

5.3.4 Boosted Multi-Variate Analysis

To enhance the separation between signal and background processes, a multivariate analysis based on Boosted Decision Trees (BDTs) is employed for each leptonic channel. The training strategy follows the approach used in the previous boosted analysis, with identical hyperparameter settings and training procedure. The only modification concerns the boosted topology definition, which in this work is extended to events with $p_T^V > 250$ GeV instead of $p_T^V > 400$ GeV, as discussed in the previous section.

Separate BDT classifiers are trained for the three leptonic channels (0L, 1L, and 2L) in order to exploit the different event topologies and reconstructed objects available in each case. A two-fold cross-validation procedure is used during the training to reduce potential biases and control overtraining. In this approach, the dataset is divided into two subsets based on the event number parity, and two independent classifiers are trained and evaluated on complementary subsets of the data.

The set of input variables provided to the BDT slightly differs across the leptonic channels in order to account for the different event topologies and reconstructed objects available in each case. In general, input variables describe the kinematics of the Higgs candidate large-R jet, the reconstructed vector boson, and the global structure of the event, including additional jets that may originate from background processes such as $t\bar{t}$ production.

The variables used for each leptonic channel are summarized in Table 5.3, and are the following:

- m_J: invariant mass of the large-R jet associated with the Higgs boson candidate.
- p_{T}^J: transverse momentum of the large-R jet.
- p_{T}^V: transverse momentum of the vector boson candidate reconstructed from the leptonic system (or from missing transverse momentum in the 0-lepton channel).
- $|\Delta\phi(V, J)|$: absolute azimuthal angle difference between the vector boson candidate and the large-R jet.
- $|\Delta y(V, J)|$: absolute rapidity angle between the vector boson and the large-R jet.
- MET: missing transverse energy associated to the 0- and 1-lepton channel events 8in the 0-lepton channel corresponds to p_{T}^V).
- p_{T}^ℓ: transverse momentum of the charged leptons in the 1- and 2-lepton channels.
- $m_{\ell\ell}$: invariant mass of the dilepton system, used only in the 2-lepton channel.
- $\cos\theta_{\mathrm{lep}}$: cosine of the helicity angle between the 2 leptons in the dilepton rest frame.
- $N_{\mathrm{AdditionalCaloJets}}$: number of additional small-R calorimeter jets not overlapping with the large-R jet.
- lepPtBalance: transverse momentum balance between the lepton and the vector boson candidate, used only in the 1-lepton channel and defined as $(p_T^\ell - E_T^{\mathrm{miss}})/p_T^W$.
- $p_{\mathrm{T}}^{\mathrm{smallR},i}$: transverse momentum of the i-th leading small-R jet not overlapping with the large-R jet ($i = 1, 2, 3$).
- $\Delta R(J, \mathrm{smallR}, i)$: angular separation in the η - ϕ plane between the large-R jet and the i-th leading non-overlapping small-R jet ($i = 1, 2, 3$).

Table 5.3: Input variables used for the BDT training in each leptonic channel.

Variable	0-lepton	1-lepton	2-lepton
mJ	x	x	x
pTJ	x		
pTV	x	x	x
absDeltaPhiVJ	x	x	x
absDeltaYVJ		x	x
MET	x	x	
pTL		x	x
mLL			x
cosThetaLep			x
NAdditionalCaloJets	x	x	x
lepPtBalance		x	
pTSmallRNoOverlapJ$_i$	x	x	x
dRLargeRSmallRNoOverlapJ$_i$	x	x	x

Table 5.4: Input variables to the BDT for the boosted $VH(b\bar{b})$ analysis for each leptonic channel.

Several of these variables are common across the three channels. In particular, the invariant mass of the large-R jet (m_J) provides direct sensitivity to the Higgs boson mass peak, while the transverse momentum of the reconstructed vector boson (p_T^V) and the angular separation variables $|\Delta\phi(V, J)|$ and $\Delta y(V, J)$ characterize the kinematic correlations expected in VH production. Variables related to additional small-R jets, such as their transverse momenta and angular distances from the large-R jet, help identify event topologies typical of background processes, especially top-quark pair production.

Channel-specific variables are also included to exploit the available information in each final state: For example, the missing transverse energy (MET) plays a key role in the 0-lepton and 1-lepton channels, while the dilepton invariant mass ($m_{\ell\ell}$) and angular observables such as $\cos\theta_{\text{lep}}$ are relevant in the 2-lepton channel, where the vector boson is reconstructed from a pair of electrons or muons.

As discussed previously, the performance of the GN2X tagger were found to be dependent on the Monte Carlo production campaign, with specific issues observed for the simulation campaign corresponding to the 2018 data-taking period. For this reason, both the event selection and categorization, as well as the BDT training and evaluation, are performed using only the MC campaigns corresponding to the 2015-2016 and 2017 data-taking periods.

The BDTs are trained separately for each leptonic channel using simulated Monte Carlo samples corresponding to the $VH(H \to b\bar{b})$ signal and the main background processes, including $t\bar{t}$ production, single-top processes, W/Z boson production in association with jets, and diboson production. The training is performed using events satisfying the boosted pre-selection, which requires at least one large-R jet with $m_J > 50$ GeV and $p_T^J > 250$ GeV, together with exactly two b-tagged track-jets matched to the large-R jet and no additional b-tagged track-jets outside the Higgs candidate.

The distributions of the input variables used in the BDT training are shown in Figures 5.15, 5.16 and 5.17 respectively for 0-, 1- and 2-lepton channel. These variables highlight the kinematic differences between the $VH, H \to b\bar{b}$ signal and the dominant background processes, such as $t\bar{t}$ and V+jets production that each BDT exploits for the signal-background discrimination through the outptu BDT score.

Distributions of the BDT classifier response for both signal and background events in each leptonic channel are shown in Figure 5.18. Signal events predominantly populate the high-

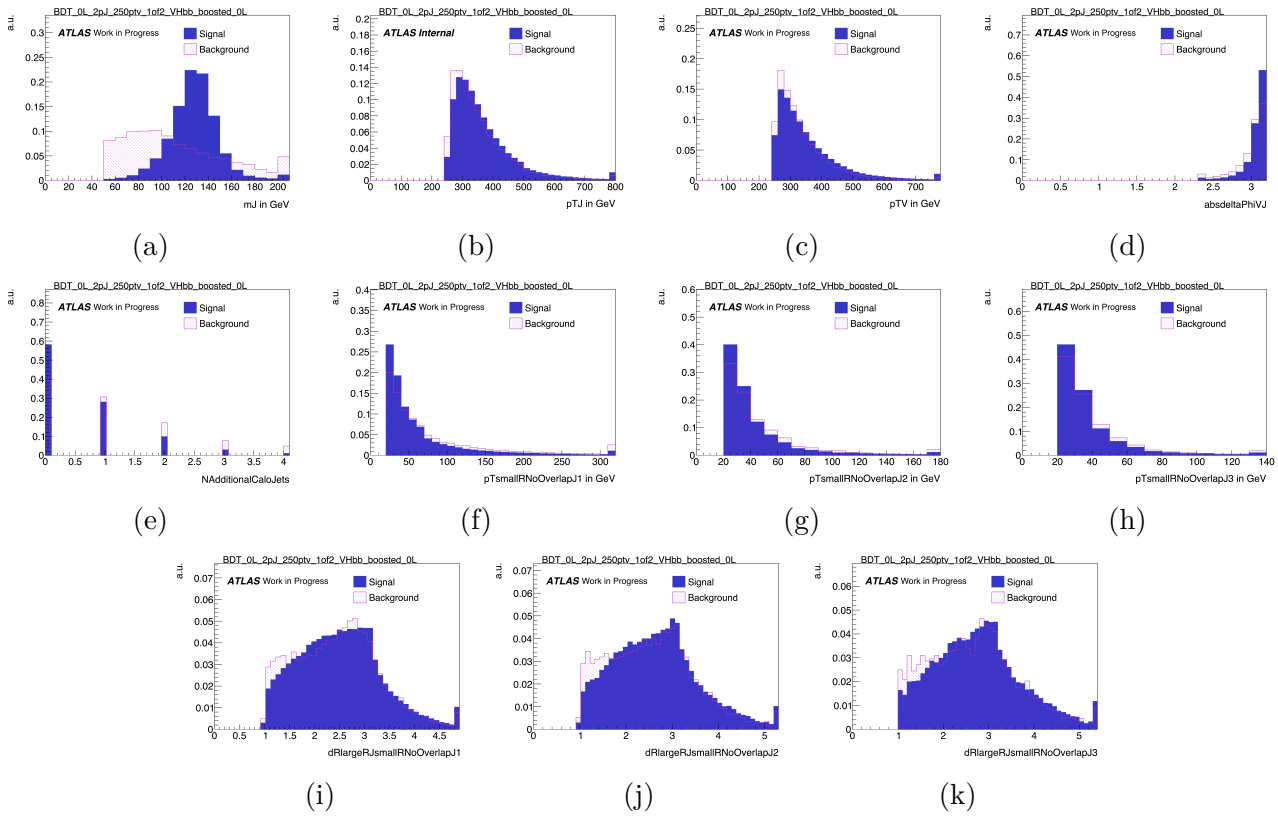


Figure 5.15: Normalized BDT input variable distributions for signal (solid blue) and background (hatched red) in 0-lepton channel for p_T^V >250 GeV. The upper bin is filled with the overflow content. Input variables are: (a) mass of the leading large-R jet, (b) p_T of large-R jet, (c) p_T^V of the vector boson, (d) $|\Delta\phi(V, H_{cand})|$ angular distance between the vector boson and the Higgs candidate, (e) number of additional small-R jets, (f,g,h) p_T of the first 3 additional small-R jets not overlapping the Higgs candidate, (i,j,k) ΔR distance between each of the 3 additional small-R jet and the Higgs candidate.

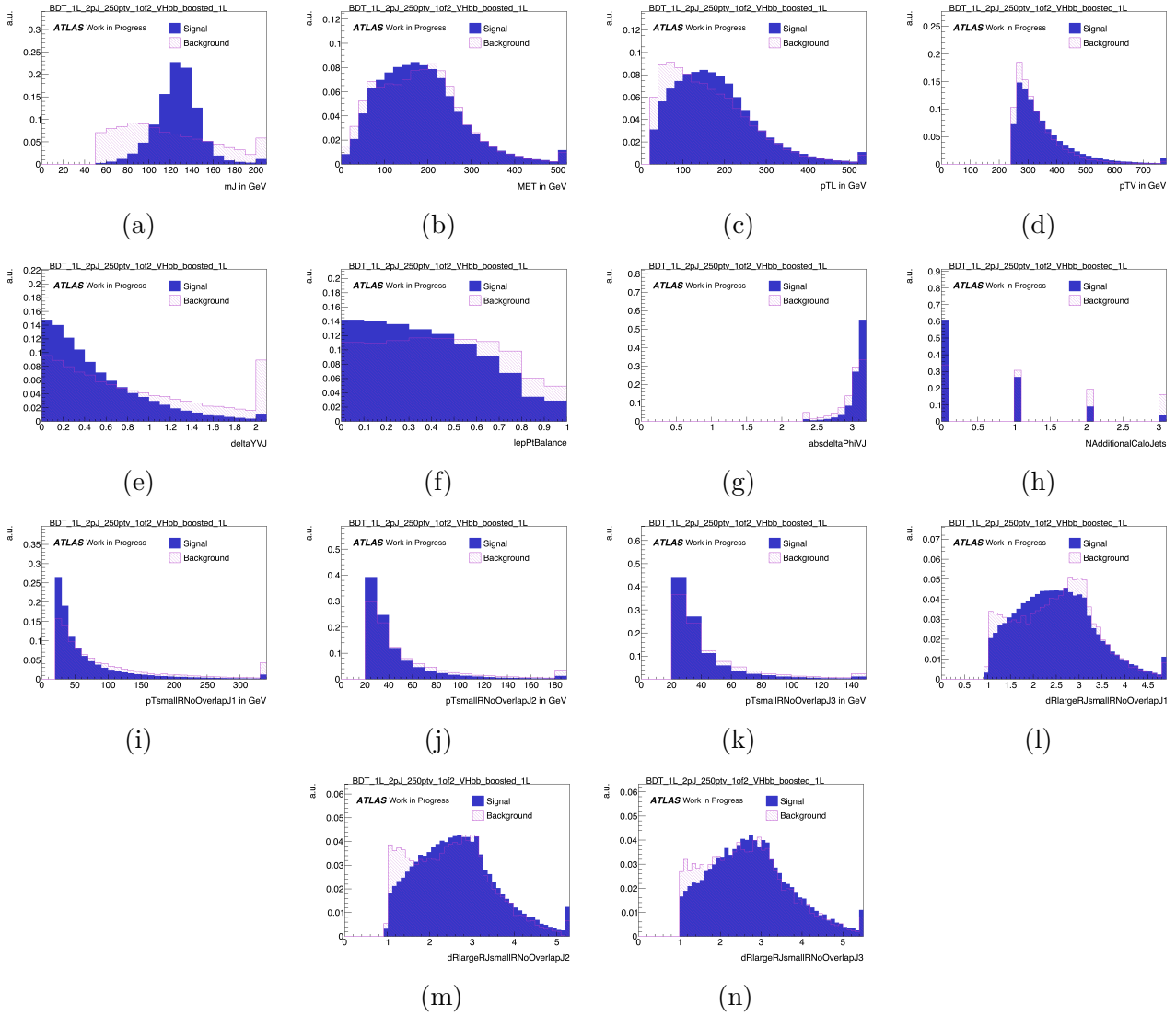


Figure 5.16: Normalized BDT input variable distributions for signal (solid blue) and background (hatched red) in 1-lepton channel for p_T^V >250 GeV. The upper bin is filled with the overflow content. Input variables are: (a) mass of the leading large-R jet, (b) missing transverse energy, (c) p_T^L of the produced lepton, (d) p_T^V of the vector boson, (e) $|\Delta y(V, H_{cand})|$ rapidity distance between the vector boson and the Higgs candidate (f) balancing variable defined as $(p_T^{\ell_1} - E_T^{\text{miss}})/p_T^W$, (g) $|\Delta\phi(V, H_{cand})|$ angular distance between the vector boson and the Higgs candidate, (h) number of additional small-R jets, (i,j,k) p_T of the first 3 additional small-R jets not overlapping the Higgs candidate, (l,m,n) ΔR distance between each of the 3 additional small-R jet and the Higgs candidate.

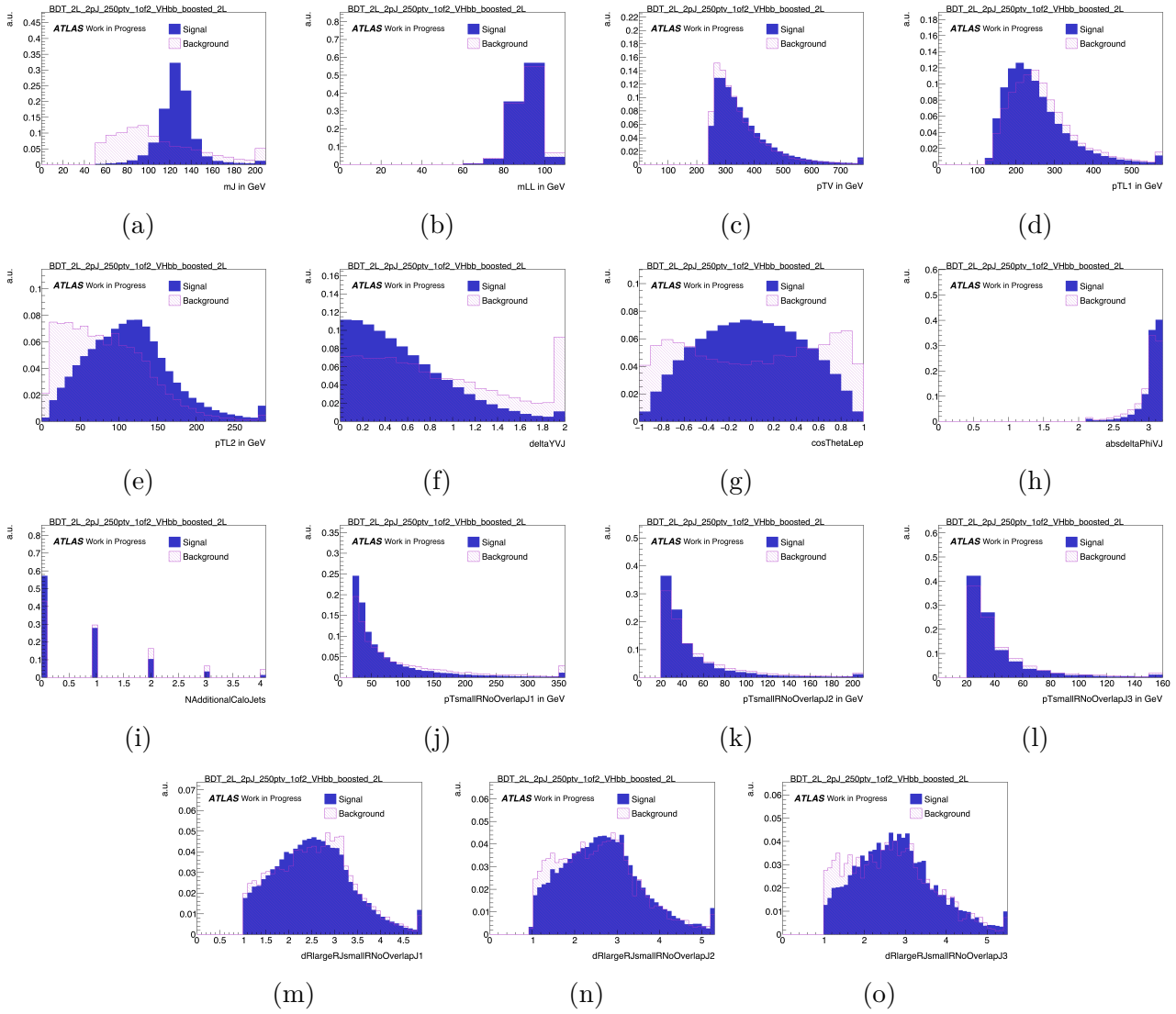
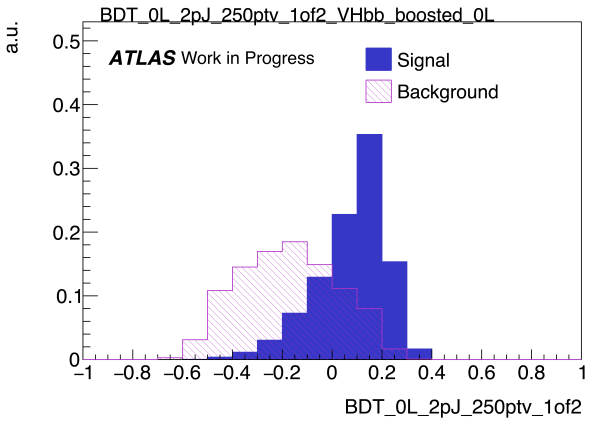
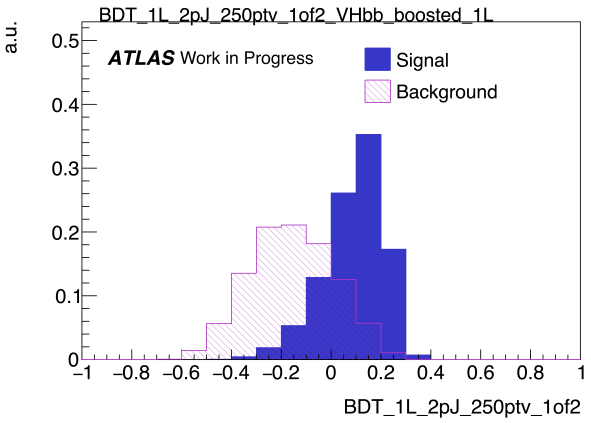


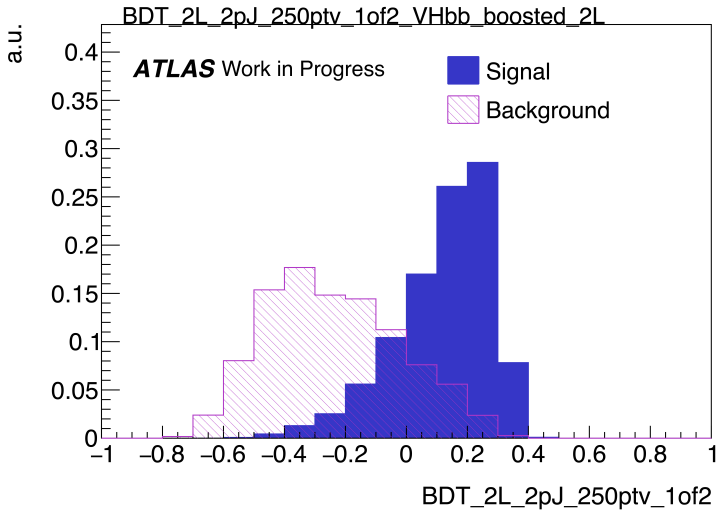
Figure 5.17: Normalized BDT input variable distributions for signal (solid blue) and background (hatched red) in 2-lepton channel for p_T^V >250 GeV. The upper bin is filled with the overflow content. Input variables are: (a) mass of the leading large-R jet, (b) m_{LL} mass of the di-lepton system, (c) p_T^V of the vector boson, (d,e) $p_{T_{1,2}}^L$ of the produced leptons, (f) $|\Delta y(V, H_{cand})|$ rapidity distance between the vector boson and the Higgs candidate, (g) $cos\theta(\ell^-, Z)$, (h) $|\Delta\phi(V, H_{cand})|$ angular distance between the vector boson and the Higgs candidate, , (i) number of additional small-R jets, (j,k,l) p_T of the first 3 additional small-R jets not overlapping the Higgs candidate, (m,n,o) ΔR distance between each of the 3 additional small-R jet and the Higgs candidate.



(a)



(b)



(c)

Figure 5.18: BDT output score distribution for signal (solid blue) and background (hatched red) for: (a) 0-lepton, (b) 1-lepton and (c) 2-lepton channels for $p_T^V > 250$ GeV.

score region, while background events accumulate at lower values of the classifier output. This separation illustrates the effectiveness of the multivariate approach in enhancing signal-to-background discrimination.

By combining multiple kinematic and topological observables into a single classifier, the BDT achieves improved separation between the $VH, H \to b\bar{b}$ signal and the dominant background processes compared to any individual variable. The BDT response is therefore used as the input observable in the statistical analysis, where signal and background contributions are extracted through a simultaneous fit across the different leptonic channels and event categories.

The performance of the trained classifiers are evaluated, for each leptonic channel, in several regions defined by the transverse momentum of the reconstructed vector boson. In particular, the analysis considers the regions $p_T^V > 250$ GeV, $p_T^V > 400$ GeV, and $p_T^V > 600$ GeV, as well as the intermediate intervals $250 < p_T^V < 400$ GeV and $400 < p_T^V < 600$ GeV.

The discriminating power of the BDT can be quantified using the area under the receiver operating characteristic (ROC) curve, computed looking at the number of selected signal events and rejected background ones for each cut value on the BDT score distribution in Figure 5.18. For the inclusive boosted region $p_T^V > 250$ GeV, the ROC integrals are:

- 0-lepton channel: ROC ≈ 0.84
- 1-lepton channel: ROC ≈ 0.87
- 2-lepton channel: ROC ≈ 0.90

These values indicate good discrimination between signal and background events, with the best performance observed in the 2-lepton channel due to the cleaner reconstruction of the Z boson and the reduced background contamination.

In addition to the ROC-based evaluation, the BDT performance were also assessed comparing the sensitivity obtained with the BDT output score distribution to the mass-based one in each evaluation region. Specifically, the expected significance, was compared between the large-R jet mass distributions and the BDT output distributions after the Transformation D.

5.3.5 Full Run-2 Significance studies

As the goal of this Chapter is to assess the potential improvement from the inclusion of the calibrated GN2X tagger in the $VH, H \to b\bar{b}$ boosted analysis, results over the full Run-2 dataset are required.

Since the studies are based on Monte Carlo samples corresponding to the 2015–2017 data-taking period, the results are extrapolated to full Run-2 by exploiting the scaling of the sensitivity with the square root of the luminosity. The obtained significances extracted from BDT output score and mass histograms are reported, for the full Run-2, in Tables 5.5, 5.6 and 5.7. The third column shows the relative improvement in percentage obtained when using the BDT response.

It can be observed that the BDT consistently provides higher significance across all p_T^V regions in all leptonic channels. The improvement is most pronounced in the 1-lepton channel, with gains up to $\sim 60\%$ in the inclusive boosted region ($p_T^V > 250$ GeV), reflecting the stronger separation power of the multivariate classifier over the single-variable mass distribution. In the 0- and 2-lepton channels, the improvement ranges roughly between 14% and 34%, with slightly larger gains in the intermediate p_T^V intervals. These results confirm that the BDT effectively enhances the signal-to-background discrimination and increases the expected sensitivity of the analysis compared to using the large-R jet mass alone.

The results obtained with the inclusion of GN2X within the boosted $VH, H \to b\bar{b}$ analysis are also compared, if selecting only events with $p_T^V > 400$ GeV, to the boosted results of the

Table 5.5: Comparison of BDT output significances and m_{bb}-based significances for the 0-lepton channel after rescaling in full Run-2, with percentage improvement.

p_T^V Region [GeV]	m_{bb} significance	BDT output significance	Improvement [%]
> 250	3.62	4.63	28
$250 - 400$	3.01	3.70	23
> 400	2.35	2.81	20
$400 - 600$	2.07	2.44	18
> 600	1.26	1.44	14

Table 5.6: Comparison of BDT output significances and m_{bb}-based significances for the 1-lepton channel after rescaling in full Run-2, with percentage improvement.

p_T^V Region [GeV]	m_{bb} significance	BDT output significance	Improvement [%]
> 250	2.97	4.74	60
$250 - 400$	2.39	3.75	57
> 400	2.12	2.99	41
$400 - 600$	1.85	2.59	40
> 600	1.39	1.65	37

Table 5.7: Comparison of BDT output significances and m_{bb}-based significances for the 2-lepton channel after rescaling in full Run-2, with percentage improvement.

p_T^V Region [GeV]	m_{bb} significance	BDT output significance	Improvement [%]
> 250	2.63	3.44	31
$250 - 400$	2.18	2.78	27
> 400	1.54	2.07	34
$400 - 600$	1.39	1.86	34
> 600	0.71	0.94	32

previous analysis, in which the DL1r tagger was used for selecting the signal events. Table 5.8 shows the inclusive significance computed for $p_T^V > 400$ GeV in the two analyses using both mass and BDT distributions after applying Transformation D. For each of the two analysis, inclusive results are obtained with a sum over each p_T^V interval and leptonic channel value, also the relative improvement achieved with the GN2X inclusion with respect to DL1r is reported.

	m_{bb} significance	BDT significance
$VH, H \to b\bar{b}$ with DL1r	2.11	3.78
$VH, H \to b\bar{b}$ with GN2X	3.63	4.67
Improvement	71.8%	23.5%

Table 5.8: BDT- and mass-based inclusive statistical only significance values when using DL1r flavour tagging algorithm (the previous Run-2 analysis [62]) and when including GN2X. Values reported are for p_T^V>400 GeV.

The results shown are purely statistical-only and were obtained with identical MVA input features and BDT training setup and hyper-parameters in both configurations; only the flavour-tagging algorithm was changed from DL1r to GN2X. Under these conditions the inclusive BDT-based significance increases from 3.78 to 4.67 ($\simeq +23.5\%$ relative) and the m_{bb}-only significance from 2.11 to 3.63 ($\simeq +71.8\%$ relative). Because the BDT inputs and training are the same, the observed gains can be attributed directly to the improved tagging performance of GN2X

rather than to changes in the multivariate classifier. The much larger relative improvement for the m_{bb}-only case indicates that GN2X especially reduces sources of background or mistag rate that degrade the dijet-mass peak. Because the BDT already combines the mass variable with several other discriminating variables, some of the benefit from a better flavour tagger is absorbed by correlations among inputs.

After having provided a fair comparison with the last ATLAS public results regarding the boosted $VH, H \to b\bar{b}$ channel [62], now a new step in this analysis using GN2X can be done. As anticipated, with respect to the latest result, the approach presented in this thesis consider the events in the 250–400 GeV region to be processed by the boosted GN2X tagger, while in the previous analysis events relying in this p_T^V bin were analyzed using the resolved strategy. The inclusion of the 250–400 GeV bin results in an overall increase in sensitivity, reaching an inclusive value of 7.52 for $p_T^V > 250$ GeV. A direct comparison of this result with the significance obtained when employing the GN2 tagger within the resolved strategy for this kinematic region is deferred to future studies and is therefore beyond the scope of the present thesis. Systematic uncertainties are not included at this stage and would need to be evaluated separately in the context of the full statistical interpretation.

Within each of these signal regions, the BDT output score is used as the final discriminant and will serves as input to the statistical fit. Some of the obtained BDT signal region distributions reported in Figures 5.19, 5.20 and 5.21 therefore represent the key ingredients entering the extraction of the signal in the presence of the GN2X-enhanced boosted selection. In Figures 5.19, 5.20 and 5.21 the blinded BDT score distributions, showing the data to Monte Carlo agreement for each of the selected p_T^V region and category, are reported, together with the expected sensitivity for each configuration.

The fit procedure itself is deferred to future studies, pending the inclusion of the full set of relevant systematic uncertainties. In the 0-lepton channel, events are divided into four categories based on the presence of additional objects beyond the $b\bar{b}$-tagged large-R jet that defines the signal region: no additional objects (`0jet0taus`), one additional τ (`0jet1ptaus`), one additional small-R jet (`1pjet0taus`), and one jet together with one τ (`1pjet1ptaus`). In the 1-lepton channel, events are separated into two categories depending on whether an additional small-R jet is present (`1pjet`) or not (`0jet`). The 2-lepton channel is not further subdivided, with the only requirement being the absence of additional associated small-R jets in the final state.

Overall, Table 5.8 clearly demonstrates that the $VH, H \to b\bar{b}$ sensitivity in the boosted regime can be significantly enhanced through the sole adoption of the GN2X-based signal event selection, even without any modification to the multivariate discriminant. In particular, an expected improvement of 23.5% in the statistical-only significance is observed with respect to the latest DL1r-based approach, highlighting the strong discriminating power of the GN2X tagger at the object-selection level. This result underscores the potential of next-generation flavour-tagging algorithms to directly impact the physics reach of the analysis, already at an early stage of the event selection chain.

Building on these findings, a natural next step is to investigate whether further gains in sensitivity can be achieved by exploiting the information provided by the GN2X tagger beyond the selection stage. To this end, the following chapter is dedicated to a set of exploratory studies in which GN2X-related observables are incorporated as additional inputs to the BDT training, with the aim of maximizing the overall discrimination between signal and background processes. These studies are intended to provide an outlook on possible future developments of the boosted $VH, H \to b\bar{b}$ analysis, and to assess the extent to which a more integrated use of advanced tagging techniques can further enhance the expected sensitivity.

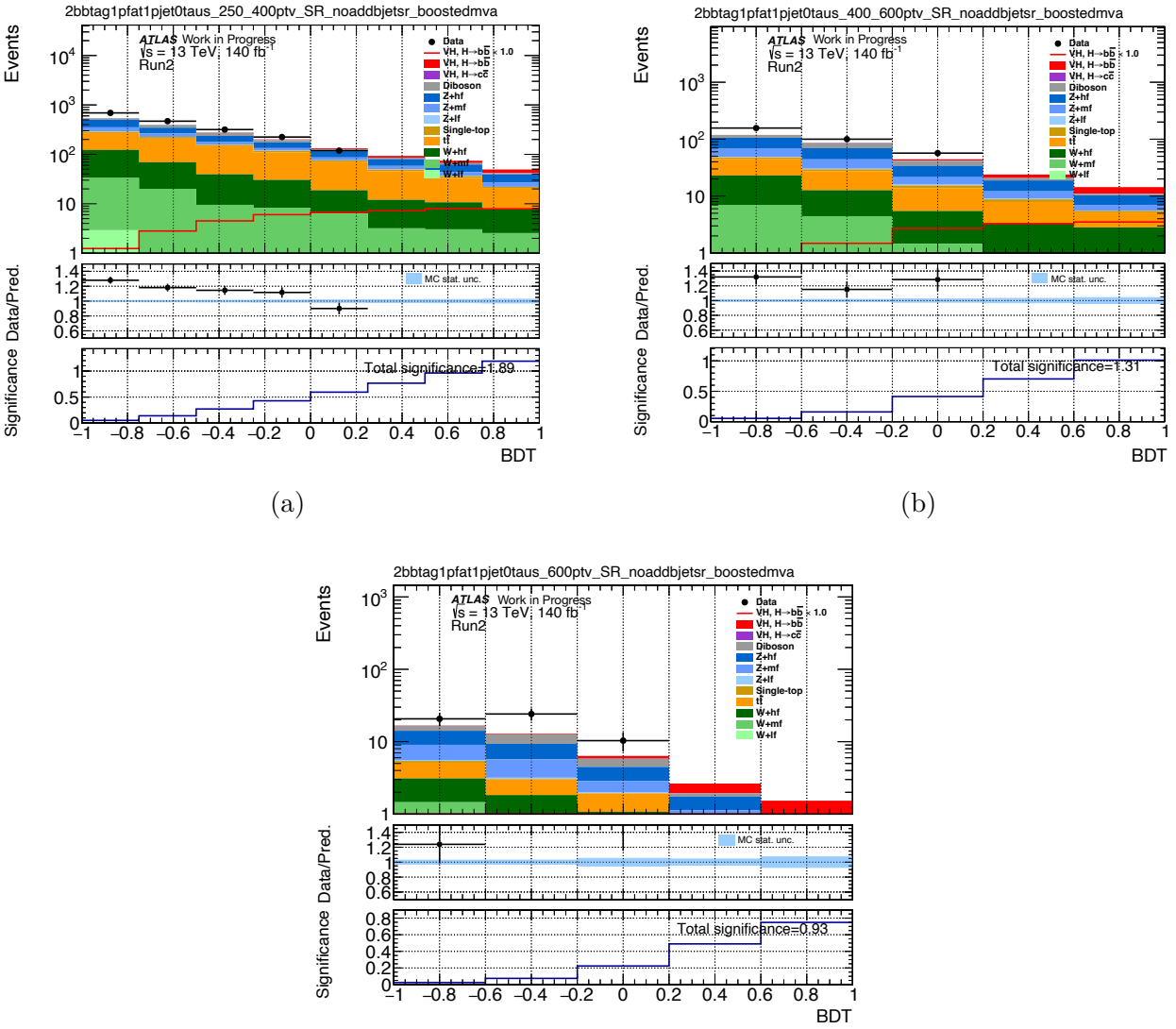
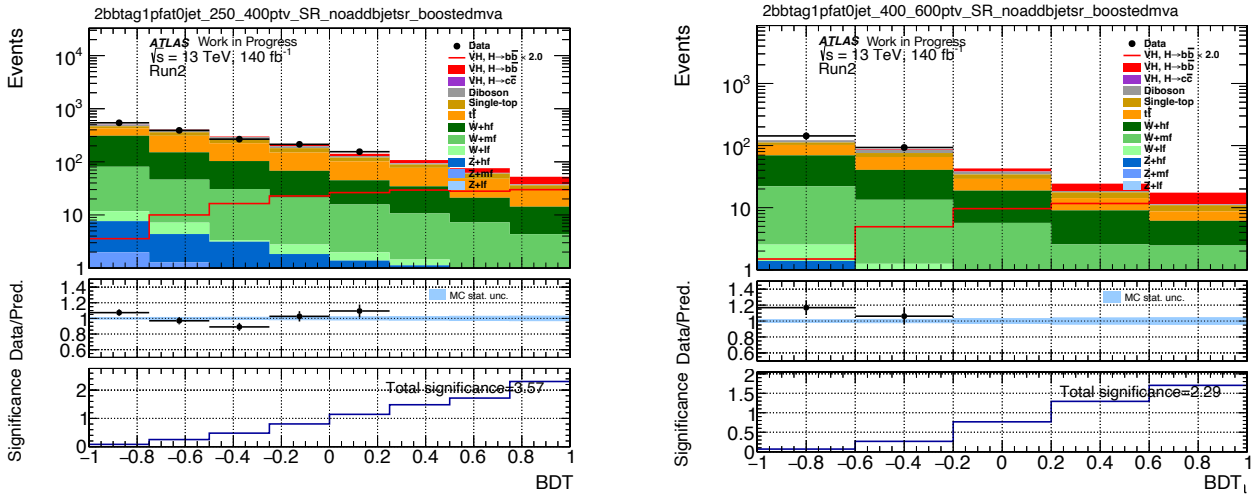
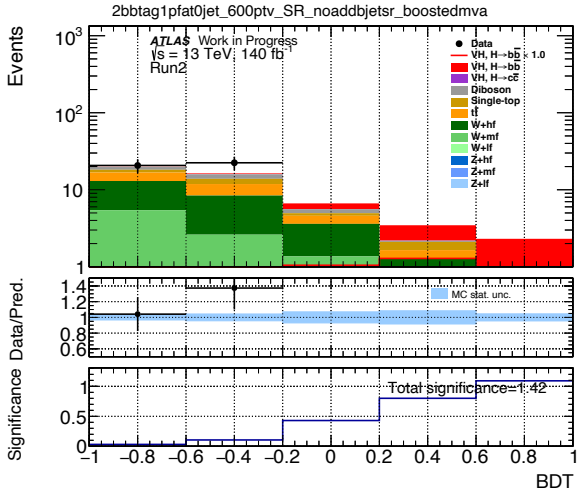


Figure 5.19: 0-lepton channel BDT score distributions for the GN2X 70% working point. BDT distributions are reported here only for the `1pfat0taus` category in the full Run-2 in (a) [250-400] GeV, (b) [400-600] GeV, and (c) >600 GeV. For each p_T^V range also blinded data points are added.

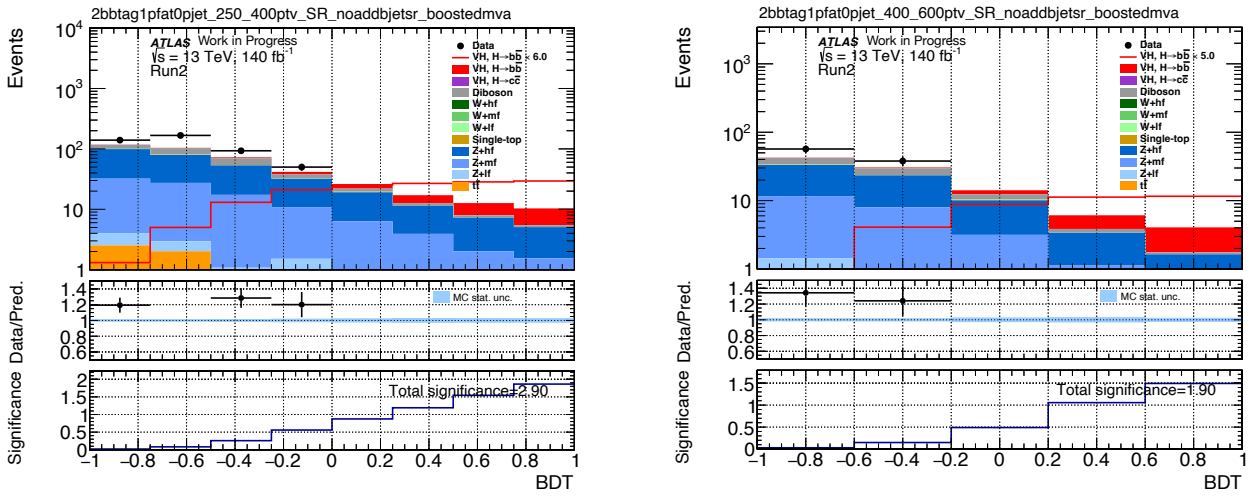


(a) (b)

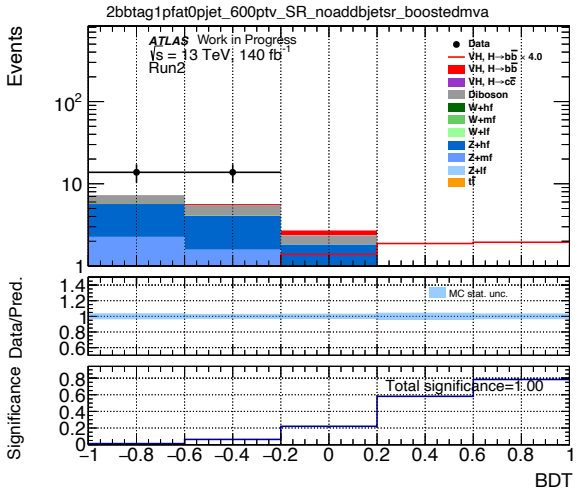


(c)

Figure 5.20: 1-lepton channel BDT score distributions for the GN2X 70% working point. BDT distributions are reported here only for the `0jet` category in the full Run-2 in (a) [250-400] GeV, (b) [400-600] GeV, and (c) >600 GeV. For each p_T^V range also blinded data points are added.



(a) (b)



(c)

Figure 5.21: 2-lepton channel BDT score distributions for the GN2X 70% working point. BDT distributions are reported here only for the full Run-2 in (a) [250-400] GeV, (b) [400-600] GeV, and (c) >600 GeV. For each p_T^V range also blinded data points are added.

5.3.6 Further studies: GN2X Tagger inclusion in MVA for boosted $VH, H \to b\bar{b}$ analysis

The results discussed in the previous section were obtained by employing the GN2X tagger within the analysis workflow. This approach already demonstrated an improvement in the final sensitivity with respect to previous-generation taggers, as reported in Table 5.8. However, this implementation does not exhaust the possible ways of exploiting the information provided by GN2X. In the context of this thesis, and with a view towards future developments of the boosted $VH, H \to b\bar{b}$ analysis, additional studies were performed to investigate alternative strategies for incorporating the GN2X output into the analysis workflow. In particular, the goal was to assess whether including GN2X-related information directly as input to the multivariate discriminants could further enhance the overall performance. As done before, due to the Monte Carlo campaign dependence of the GN2X tagger, all the results are obtained training only on simulations corresponding to the 2015–2017 data taking period. Also, as these studies were performed before obtaining the Scale Factor values for GN2X, these are not included in the following results.

Following this strategy, an initial study was carried out by adding to the standard set of BDT input variables reported in Table 5.3, the kinematic information related to the small-R jets contained within the reconstructing large-R jet, excluding the jet mass. Specifically, for the first three reconstructed small-R jets, the transverse momentum of each of them, together with the angular distances ΔR between the selected small-R jets and between each of them and the large-R jet, were included. This choice was motivated by the observation that such information is, in principle, already exploited by the GN2X tagger and encoded in its output score used for selecting signal events. The study therefore aimed to test whether this information is fully captured by the GN2X score alone, or whether providing it explicitly to the BDT could lead to an additional gain in discrimination.

The results of these studies are presented in Fig. 5.22 for the 0-lepton channel. The statistical-only significance is compared across three configurations: a mass-only approach, where the significance is computed using the leading large-R jet mass distribution; a baseline BDT using the BDT output score obtained when training with standard input variables described in the Table 5.3; and an extended BDT-based result including, in addition, the small-R jet-related variables.

Obtained results show that, with the inclusion of small-R jets related variables as input, the largest improvement is observed for the 50% WP. Overall, for each working point and evaluation region, an improvement is observed both in the standard BDT sensitivity and when incorporating small-R jet information as additional input.

For the results presented in the following, the focus is placed on the 70% working point. This choice is primarily motivated by the fact that the GN2X tagger calibration performed in this thesis was carried out at this working point, which was subsequently adopted in the boosted $VH, H \to b\bar{b}$ analysis, as described in the previous sections of this chapter. As such, it represents the most relevant and consistent configuration for a detailed study within the analysis framework. It should be emphasized, however, that all the studies discussed in this section were performed for the full set of available working points.

A second set of studies was performed with the goal of probing the maximum achievable improvement in sensitivity when exploiting the information provided by the GN2X tagger. In this case, the continuous GN2X score was included directly as an input variable to the BDT training, in addition to the small-R jet-related variables described above. The motivation for this approach is that the continuous GN2X response represents the most complete and fine-grained encoding of the tagger information at the event level. Providing this quantity directly to the multivariate discriminant allows one to estimate the maximum achievable discrimina-

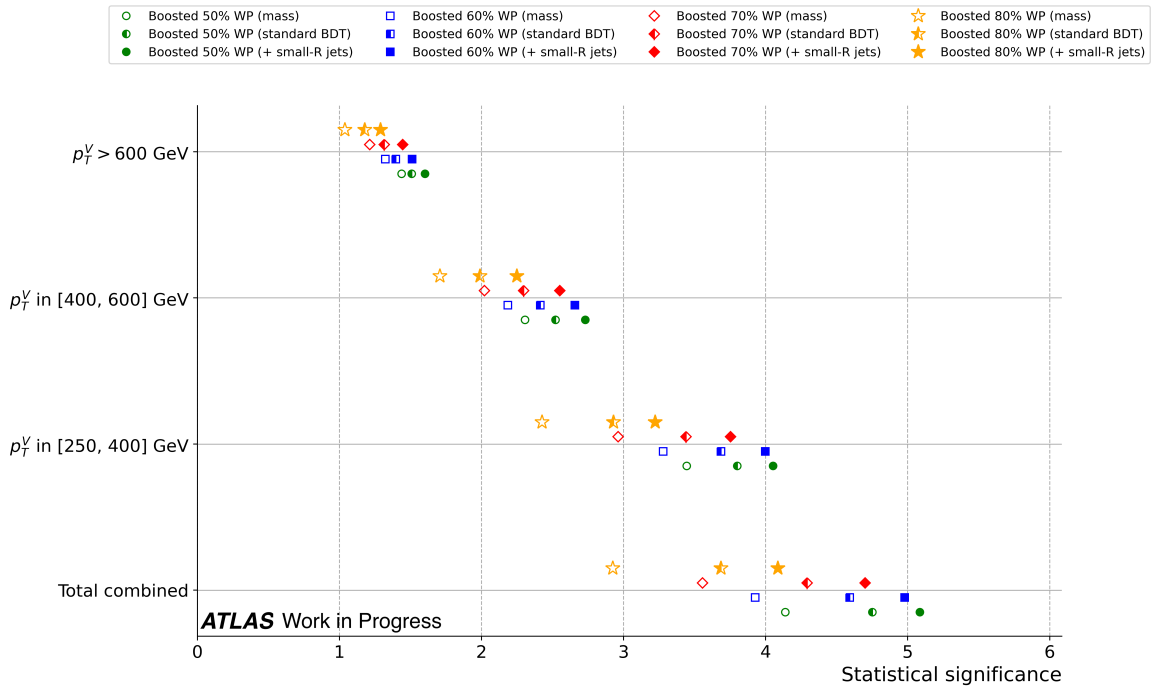


Figure 5.22: Statistical-only significance in the 0-lepton channel for different p_T^V regions and for the total combination. Results are shown for four GN2X working points (50%, 60%, 70%, and 80%), comparing three analysis strategies: a mass-only approach based on the leading large-R jet mass (mass), a BDT using the standard input variables (standard BDT), and a BDT including additional kinematic variables related to small-R jets within the large-R jet (+ small-R jets).

tion performance that can be reached. At the same time, it is important to stress that this configuration is intended as a performance benchmark rather than a realistic analysis strategy. The continuous GN2X score cannot be directly used in physics measurements, as its continuous response cannot be calibrated in a way that would allow for its inclusion in physics analyses.

The results are presented in Fig. 5.23. In all configurations, the GN2X tagger is applied at the 70% signal efficiency working point exclusively at the pre-selection stage, where it is used in conjunction with the kinematic requirements to reconstruct the signal candidates. Building on this common pre-selection, the sensitivity is then evaluated and compared across different analysis strategies: a mass-based approach relying on the leading large-R jet mass, a BDT trained with the standard set of input variables, an extended BDT including additional small-R jet-related information, and finally a configuration where the continuous GN2X score is also provided as input to the training. A clear improvement in sensitivity is observed when the continuous GN2X score is included among the BDT input variables. In all p_T^V regions, as well as in the total combination, this configuration yields the highest statistical significance. This behavior highlights that the continuous GN2X response carries additional discriminating information that is not fully exploited when using only working point-based selections. This result represents an estimate of the maximal sensitivity improvement that can be achieved when fully exploiting the GN2X information.

It is important to stress, however, that this configuration is not directly applicable in a realistic analysis due to the continuous behaviour of the included GN2X information. For this reason, additional studies were carried out to investigate how this information could be incorporated in a way that can be then calibrated, while preserving as much as possible of the observed improvement. To this end, a pseudo-continuous approach was explored, in which the GN2X output is discretized into quantile-based bins, and each event is assigned to a bin

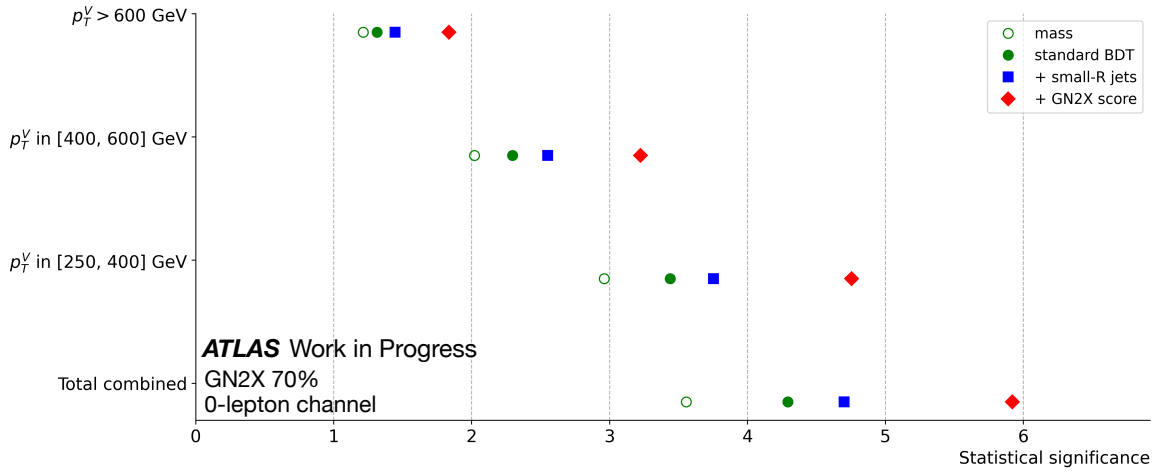


Figure 5.23: Statistical significance obtained for the 70% GN2X working point in the 0-lepton channel with different training setup: when using leading large-R jet mass distribution (mass), when using the standard input set of variables and compute significance from BDT output score (standard MDT), when adding to the standard set p_T and relative ΔR information regarding small-R jets within the reconstructed large-R jet (+ small-R jets) and when adding for each event the corresponding GN2X discriminant output value (+GN2X score). The continuous implementation of the GN2X tagger as input lead to the maximum reachable performance with tagger inclusion in the BDT.

according to its corresponding score value.

Different binning granularities were considered, corresponding to bin widths of 5%, 10%, and 15%, as well as a minimal configuration with only two bins. In particular, considering 70% GN2X selection applied, for each configuration the obtained bin-content are:

- **5% bin-width**: 70-65 %, 65-60 %, 60-55 %, 55-50 %, <50 %;
- **10% bin-width**: 70-60 %, 60-50 %, <50 %;
- **15% bin-width**: 70-65 %, 65-50 %, <50 %;
- **2 bins**: 80-50 %, <50 %.

In this way, a pseudo-continuous representation of the GN2X response is constructed and used as an additional input variable to the BDT training, allowing one to retain a significant fraction of the original information while remaining compatible with calibration procedures [68].

In Figure 5.24, the signal and background distributions of this variable are shown for the different binning configurations. These distributions are obtained after the 70% GN2X application in the events pre-selection stage, mapping the selected events in one of the defined quantile-bins as a function of their GN2X score.

The statistical significance obtained for the different bin-width configurations is reported in Table 5.9. The results show that the sensitivity remains remarkably stable across all considered binning schemes. In all cases, the achieved significance is lower than that obtained with the continuous implementation of the GN2X score, confirming that some information is inevitably lost through discretization. At the same time, the observed stability indicates that increasing

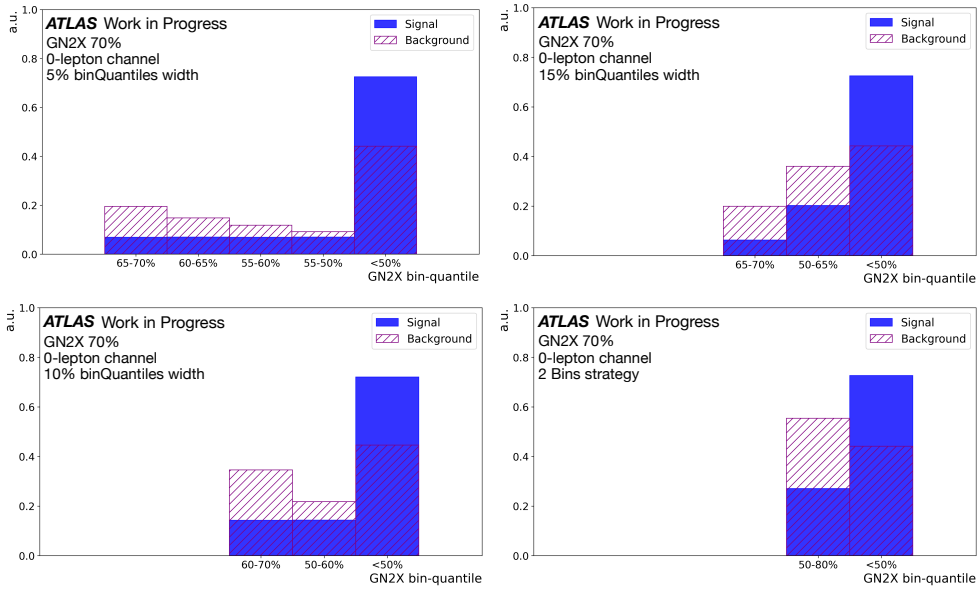


Figure 5.24: GN2X quantized information input distribution for both signal and background in the 0-lepton channel for events passing the 70% working point. Results for four different configuration of the bin width are showed: 5% bin width (upper left), 10% bin width (lower left), 15% bin width (upper right) and the 2 bins strategy (lower right).

GN2X Info:	5% bin-width	10% bin-width	15% bin-width	2-Bins strategy	Continuous
Significance:	5.35	5.35	5.32	5.33	5.92

Table 5.9: Statistical significance obtained with different quantized inclusions of the GN2X information as input variable. Results obtained different configuration of the bin width are presented here together with the significance obtained with the continuous information of the GN2X score as input to the BDT training. The results are obtained adding for each configuration, together with the standard BDT input variables, also the small-R jets information.

the bin granularity does not lead to a substantial performance gain. This opens the possibility of adopting a simplified approach for future studies. In particular, the two-bin configuration emerges as a promising option, as it provides comparable sensitivity while offering a significantly simpler calibration procedure with respect to finer binning schemes.

In Table 5.10, the statistical-only sensitivity at the 70% GN2X working point is reported and compared between two scenarios: using the standard set of input variables with the tagger information implemented via bin quantiles, and augmenting this with information from the small-R jets contained within the reconstructed large-R jet. For all bin-width configurations, the results show a consistent decrease in performance when small-R jet information is excluded. This indicates that GN2X information alone, even when quantized, cannot fully capture all the relevant features of the large-R jet. Including small-R jet information explicitly, alongside the quantized GN2X inputs, systematically improves the model's performance, and will be considered for future training studies.

A summary plot for the 70% working point is reported in Figure 5.25, in which, in addition to the significance values reported in Figure 5.23, also the values obtained with the quantized inclusion of GN2X as input variable to the BDT is reported. In particular, as each quantization scheme of the discriminant result to have similar results, the 2-bin strategy was chosen and reported into the plot.

For each evaluation region, the significance values show that the performance improves

Quantile binning	Standard MVA	MVA + small-R jets	Improvement
5%	4.97	5.35	-7%
10%	4.97	5.35	-7%
15%	4.98	5.32	-6.4%
2-bins	4.97	5.33	-7%

Table 5.10: Significance values obtained using the standard set of input variables (Standard MVA) and including small-R jet information (MVA + small-R jets), together with the percentage decrease in performance when small-R jets are excluded. Results are shown for each bin-quantile configuration at the 70% GN2X working point in the 0-lepton channel.

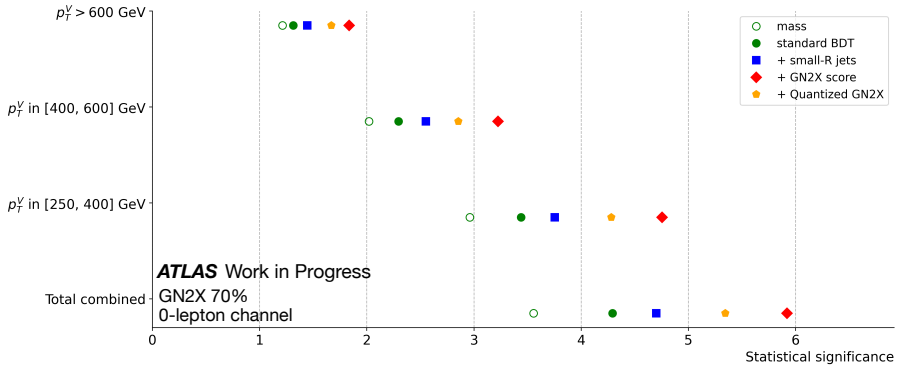


Figure 5.25: Statistical significance obtained with the 70% GN2X working point in the 0-lepton channel with different strategies: using leading large-R jet mass distribution (mass), using the BDT output score when training with the standard set of input variables (standard BDT), when adding to this set the small-R jet information (+small-R jets), when adding to this set the continuous information of the GN2X discriminant score for each event (+GN2X score) and when adding to the standard+small-R jet set of input variables the 2-bin quantized GN2X information.

when additional information is included in the BDT training inputs, first with the inclusion of small-R jet variables, and further with the addition of the quantized GN2X information. The red rhombus markers, corresponding to the inclusion of the continuous GN2X output discriminant on top of the small-R jet information, represent the maximum achievable performance in each evaluation region when tagger information is fully exploited. These results are promising for future analyses, as they demonstrate that incorporating a calibrated tagger within the BDT input set variables enhances the discrimination power, and consequently the significance, thereby opening the possibility for further improvements to be explored in future iterations of the analysis.

Chapter 6

An End-to-End fully differentiable approach to tracking

6.1 Introduction

In recent years, machine learning (ML) techniques have begun to significantly transform the way event reconstruction is performed in high-energy physics experiments. Rather than relying exclusively on handcrafted features derived from high-level detector objects, modern ML approaches increasingly operate directly on lower-level detector information, allowing algorithms to learn complex correlations directly from the raw data. A clear example of this paradigm shift is provided by recent developments in jet flavour tagging, where transformer-based neural networks achieve improved performance by exploiting the detailed structure of tracks associated with jets, without relying explicitly on reconstructed secondary vertices.

This move toward low-level, data-driven reconstruction is not limited to jet tagging. Similar approaches are now being explored across the full reconstruction chain, including particle tracking, where advanced machine learning architectures such as graph neural networks and transformer-based models have recently been proposed to address the combinatorial complexity of track reconstruction. In these approaches, tracking can be formulated as a pattern recognition problem on graphs or sequences of detector hits, allowing the relationships between hits to be learned directly from data.

Despite their promising performance, the application of deep neural networks to fundamental reconstruction tasks raises important conceptual challenges. Traditional reconstruction algorithms often incorporate well-understood physical principles through explicit inductive biases, such as constraints derived from detector geometry or the expected motion of charged particles in a magnetic field. When neural networks are used as largely black-box models, it is not always guaranteed that these physically motivated structures are naturally learned during training.

One of the most critical components of any high-energy physics experiment where these considerations become particularly relevant is charged-particle tracking. Tracking consists of reconstructing particle trajectories and estimating their momenta from the hits left in the detector, and it plays a central role in both offline reconstruction and online event selection. The performance of higher-level reconstructed objects, as well as the effectiveness of trigger strategies, strongly depends on the accuracy and efficiency of the tracking algorithms.

Standard tracking algorithms traditionally rely on a factorized reconstruction approach. Starting from the raw detector hits, reconstruction proceeds by first forming tracklets, i.e. short track segments built by locally combining spatially compatible hits in adjacent detector layers under simple geometric constraints. These tracklets provide initial direction and curvature estimates and are subsequently extended into track candidates through combinatorial pattern

recognition, where additional hits are iteratively associated according to compatibility criteria based on detector resolution and magnetic-field constraints.

In the Muon Spectrometer, track reconstruction follows a similar strategy but employs dedicated pattern-recognition techniques adapted to the detector geometry. In particular, candidate muon trajectories are first identified using a Hough transform–based algorithm [69], which searches for aligned hit patterns across the muon chambers. The resulting patterns are used to build local track segments that are subsequently combined across multiple detector layers to form full muon track candidates. These candidates are then refined through dedicated fitting procedures that account for the magnetic field and detector resolution, yielding the final estimate of the muon trajectory and momentum.

Each of these steps is optimized independently, typically using dedicated algorithms and hand-tuned criteria.

In recent years, the ATLAS collaboration has explored the use of machine learning techniques to improve tracking performance, including the adoption of recurrent neural networks and other deep learning models for specific subtasks [70]. These approaches have demonstrated promising gains in efficiency and execution time. However, even in these cases, the tracking problem has largely remained factorized, with different machine learning models trained independently for different stages of the reconstruction chain.

At the same time, recent developments in differentiable programming have enabled the integration of physics-informed constraints directly into the training of machine learning models ([71], [72], [73]). By embedding known physical relationships into differentiable architectures, it becomes possible to guide the optimization process using both data-driven learning and domain-specific knowledge. This paradigm has been shown to improve training stability, interpretability, and overall performance in a variety of scientific applications [74].

Motivated by these advances, we explored the use of differentiable programming techniques to address the tracking problem in an end-to-end unique model. The goal was to construct a single machine learning model in which all stages of the tracking procedure —hit classification, track identification, and momentum estimation— are defined within a unified, fully differentiable framework. In this approach, the entire reconstruction chain is described by a common set of parameters and optimized jointly with respect to the final task, rather than treating each subtask independently.

As a proof of concept, this methodology was applied to simulated events in a detector geometry designed to emulate the ATLAS muon spectrometer, containing a single muon track per event. Starting directly from the detector hits, the model identifies the muon trajectory and performs a circular fit to reconstruct the muon p_T.

Although the study is limited to a simplified scenario, it demonstrates the potential of end-to-end differentiable tracking as a viable alternative to traditional factorized approaches. Importantly, the proposed framework is not intrinsically tied to a specific detector geometry or experimental setup. The underlying idea can be generalized to more complex detectors, opening the possibility for future applications in both offline reconstruction and online triggering systems.

The remainder of this chapter is organized as follows. First, the simulated dataset used for this study is described, including the detector geometry, the event generation, and the characteristics of the recorded hits. Subsequently, an overview of graph-based neural networks is provided, with particular emphasis on Graph Attention Networks, which form the core building block of the proposed approach. The chapter then introduces the end-to-end differentiable tracking model in detail, describing its architecture, the incorporation of physics-informed differentiable components, and the training strategy. Finally, the performance of the model is evaluated and discussed, highlighting its ability to reconstruct single-muon tracks and estimate the transverse momentum directly from detector hits, and comparing its behavior to more

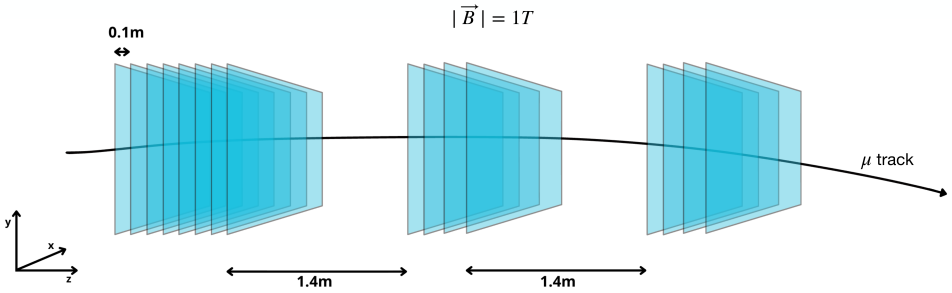


Figure 6.1: Simulated detector geometry: 16 tracking layers arranged into three stations (8+4+4) divided by 1.4 meters. The entire detector is places in a uniform magnetic field of 1 T [75].

traditional reconstruction approaches.

The complete study, including a detailed description of the model, the training procedure, and the full set of results, has been published in Ref. [75].

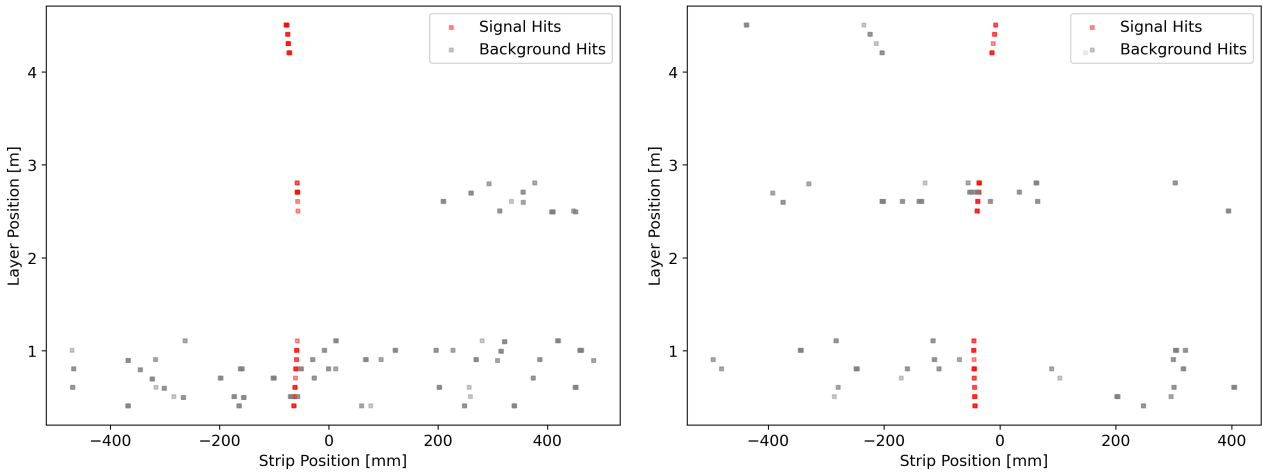
6.2 Dataset and Simulation

To study the application of differentiable machine learning techniques to track-level reconstruction, a simplified but realistic simulation of the ATLAS Muon Spectrometer was developed. In particular the input sample consists in a Monte Carlo based simulation of a generic tracking detector in which a single charged particle is present event by event.

The detector layout, depicted in figure 6.1, consists of 16 tracking layers arranged into three stations: the first station contains 8 layers, while the other two stations include 4 layers each. The entire detector is placed in a uniform magnetic field of 1 T. The simulation is done using GEANT4, with each tracking layer modeled as two printed circuit boards (PCBs) 1 cm thick, separated by a sensitive gap of 5 mm. Although this layout is intentionally generic, it draws inspiration from micro-pattern gaseous detectors used in LHC experiments, such as MicroMegas or the GEM detectors, and in particular by the design of the Phase-I ATLAS muon upgrade [76].

The detector response is simulated by extrapolating particle trajectories to the readout planes and recording the corresponding hit positions. Around each hit position, a cluster of firing strips is formed, with an average cluster size of four strips. The charge collected on each strip is sampled from a Gaussian distribution centered on the hit position, with a standard deviation equal to two-thirds of the strip pitch. This ensures that nearly all of the total cluster charge remains contained within the cluster. When multiple hits produce overlapping clusters, the charges from all contributing clusters are summed channel by channel in the overlap region. This simulation framework provides a flexible yet realistic testbed for evaluating differentiable reconstruction algorithms, enabling a systematic study of their ability to recover track parameters and classify hits under controlled yet physically meaningful conditions.

Single-muon events are simulated with transverse momentum in the range $[20, 50]$ GeV. For each particle, both digitized and true hit positions are recorded, along with the corresponding



(a) (b)

Figure 6.2: Two examples of event displays, both showing the track originating from the muon particle and additional activity related to both detector noise and secondary tracks [75].

truth PDG identification number for each hit and the muon's p_T. Examples of event displays, showing hits from signal tracks as well as contributions from noise or additional scattering, are shown in Figure 6.2.

6.3 Differentiable End-to-End Model for Tracking

To address the tracking problem in a fully differentiable, end-to-end fashion, the entire model was implemented using `JAX`-based libraries [77]. In particular, the graph-structured inputs are defined through `Jraph`, the neural network architecture is implemented with Haiku, and the optimization procedure relies on `Optax` [78].

The adoption of `JAX` is not merely motivated by computational performance, but by its functional programming paradigm and native support for composable transformations. These features make it possible to differentiate through complex computational graphs that include message-passing operations and physics-inspired transformations. In this work, the reconstruction chain is therefore expressed as a single differentiable mapping from detector hits to high-level track parameters.

By construction, gradients can be propagated through every stage of the model, ensuring that both neural network components and embedded physics-motivated operations remain within a unified optimization framework. This enables the formulation of tracking as a single global learning problem, where all trainable parameters are optimized simultaneously.

Such an approach differs fundamentally from traditional multi-stage reconstruction pipelines, in which pattern recognition, track fitting, and parameter estimation are treated as sequential and partially independent steps, often involving discrete or non-differentiable procedures. Here, the entire tracking workflow, from hit-level classification to momentum reconstruction, is embedded within a coherent differentiable architecture. During training, gradient information flows across the full chain of operations, allowing the optimization process to account for correlations between subtasks rather than addressing them in isolation.

From a conceptual standpoint, the model can be factorized into three main functional blocks:

- a **graph attention network for classification**, which assigns to each detector hit a probability of being signal or background;

- a **differentiable clustering module**, which aggregates hits layer by layer using the predicted classifications;
- a **differentiable fitting module**, which performs a weighted circular fit to extract the track curvature and, consequently, the transverse momentum.

In the following section, a general overview of Graph Attention Networks (GATs) is first presented. Subsequently, the representation of single-track events as graphs within the Jraph library is described. Building on this foundation, a detailed explanation of the implemented model is provided. This includes, first, the graph-based architecture used for hit classification, followed by a description of the two subsequent differentiable modules for clustering and track fitting.

The training setup adopted for the model is then outlined. Finally, the results are presented, with particular emphasis on both the hit-classification performance and the transverse momentum reconstruction achieved by the proposed approach.

6.4 Graph Attention Networks: Principles and Formulation

Many problems in physics and machine learning involve data that are naturally represented as *graphs*, i.e. collections of objects and the relationships between them. In this representation, objects are modeled as *nodes* and their interactions as *edges*. Graph Neural Networks (GNNs) are a family of neural networks specifically designed to operate on such data structures. Unlike standard neural networks, which assume inputs arranged in regular grids (such as images or sequences), GNNs can process irregular and variable connectivity patterns.

The core idea behind GNNs is *message passing*: each node updates its internal representation by aggregating information from its neighboring nodes. By stacking multiple layers, nodes progressively incorporate information from larger portions of the graph.

Graph Attention Networks (GATs) are a particular class of GNNs that enhance this message passing mechanism by introducing an *attention mechanism*. Instead of treating all the connected nodes equally (for example through simple averaging), GATs learn to assign different importance weights to different nodes. These weights are computed dynamically from the node features themselves, allowing the model to focus on the most relevant connections in the graph.

GATs were originally proposed in Ref. [79] as an extension of graph convolutional approaches, with the goal of increasing model expressiveness and adaptability in graphs characterized by complex or irregular connectivity patterns.
Formally, consider a graph $G = (V, E)$ with $V = 1, .., N$ set of nodes and $E \subseteq V \times V$ set of edges. Each node $i \in V$ is associated with a feature vector $\mathbf{x}_i \in \Re^F$. The goal of a GAT layer is to compute updated node representations by aggregating information from all the connected nodes using learned and data-dependent attention weights, rather than relying on a fixed and uniform averaging scheme typically used in standard GNNs.

Node Feature Embedding Each node feature vector is first mapped into a latent representation space through a learnable function:

$$\mathbf{q}_i = \phi(\mathbf{x}_i) \in \Re^{F'}, \tag{6.1}$$

where ϕ is typically implemented as a linear transformation of a multilayer perceptron (MLP), and F' denotes the latent feature dimension. This embedding step allows the network to

transform raw input features into representations that are more suitable for relational reasoning on the graph.

Attention Score Computation For each direct edge $(i, j) \in E$, an unnormalized [1] attention score vector

$$\mathbf{e}_{ij} \in \Re^{F'} \tag{6.2}$$

is computed as a learnable function of the latent representations of the sender and receiver nodes:

$$\mathbf{e}_{ij} = \psi(\mathbf{q}_i, \mathbf{q}_j), \tag{6.3}$$

where ψ is tipically implemented as an MLP acting on the concatenation of $\mathbf{q}_i$ and $\mathbf{q}_j$. Each component $e_{ij}^{(d)}$ of the vector $\mathbf{e}_{ij}$ represents the unnormalized relevance of node i to node j in the $d-$th latent dimension.

Attention Normalization The attention scores are normalized across the neighborhood of each receiving node j. For each latent dimension $d = 1, .., F'$, the normalized attention coefficients are defined as

$$\alpha_{ij}^{(d)} = \frac{\exp(e_{ij}^{(d)})}{\sum_{\kappa in N(j)} \exp(e_{kj}^{(d)})}, \tag{6.4}$$

where $N(j) = \kappa | (\kappa, j) \in E$ denotes the set of nodes connected to the node j. The resulting attention vector

$$\boldsymbol{\alpha}_{ij} = (\alpha_{ij}^{(1)}, ..\alpha_{ij}^{(F')}) \tag{6.5}$$

assigns a relative importance to node i when updating node j, separately for each latent feature dimension.

Message Aggregation The updated representation of node j is obtained by aggregating the latent representations of its neighbors, weighted by the normalized attention coefficients:

$$\mathbf{h}_j = \sum_{i \in N(j)} \boldsymbol{\alpha}_{ij} \odot \mathbf{q}_i, \tag{6.6}$$

where $\odot$ denotes the element-wise (Hadamard) product. This aggregation mechanism allows the network to selectively propagate information across the graph, emphasizing nodes that are most relevant according to the learned attention weights.

Node Update Optionally, the aggregated representation $\mathbf{h}_j$ is passed through a non-linear transformation:

$$\mathbf{y}_j = \sigma(\rho(\mathbf{h}_j)), \tag{6.7}$$

where ρ is a learnable function (e.g. an MLP)= and σ is an activation function. The vectors $\mathbf{y}_j$ constitute the output node features of the GAT layer and can be fed into subsequent graph attention layers or task-specific heads.

[1] The attention scores are referred to as *unnormalized* because they are raw, unconstrained compatibility values computed independently for each edge. They do not define relative importances until a normalization step, which is typically a softmax over the neighborhood of each node, is applied, at which point they become normalized attention coefficients.

6.5 Graph Representation of Events

The decision to adopt graph neural networks was driven by the need to effectively capture correlations between detector hits. Each input event is therefore represented as a graph. In this representation, each node corresponds to a detector hit, characterized by features encoding its digitized position in the $yz-$plane. Each event is modeled as a fully connected graph, allowing information to flow freely between all hits during message passing.

A corresponding truth graph is associated to each input graph. The truth graph shares the same node and edge structure, but includes as node features the particle identifier (1 for muon and 0 for background hits) of each hit, and as a global feature the true transverse momentum of the muon in the event. As a result, each event is encoded as a `GraphTuple`, which serves as the fundamental data structure for both training and inference.

For training efficiency, graphs are grouped into batches. Since JAX requires static shapes for optimized execution on GPUs, graphs within the same batch are padded to a common size. Therefore, for each batch containing graphs with varying numbers of nodes, artificial nodes are added so that all graphs are padded to match the size of the largest graph in the batch. These additional nodes are assigned unphysical but well-defined values and are explicitly excluded from contributing to the loss function. Self-edges are added to each node to ensure that, during message aggregation, each node can explicitly retain and update its own information.

6.6 GAT-Based Classification Network

The differentiable end-to-end model implemented is based on two stacked neural network inspired by the GAT architecture. The node embedding is computed via an MLP which maps the input features $\mathbf{x}_i$ of each node i into an hidden representation of dimension 512:

$$\mathbf{q}_i = \texttt{attention\_query}(\mathbf{x}_i) \in \mathbb{R}^{512}\,. \tag{6.8}$$

The GAT attention logit is computed as an edge feature

$$\mathbf{e}_{ij} = \texttt{attention\_logit}(\mathbf{q}_i, \mathbf{q}_j) \in \mathbb{R}^{512}\,, \tag{6.9}$$

A dimension-wise softmax normalization is applied over all incoming edges of each receiving node:

$$\alpha_{ij}^{(d)} = \frac{\exp(e_{ij}^{(d)})}{\sum_{k\in\mathcal{N}(j)} \exp(e_{kj}^{(d)})}, \quad d = 1, \ldots, 512\,, \tag{6.10}$$

where $\mathcal{N}(j)$ denotes the neighbors of node j. The updated node features are computed via element-wise multiplication and aggregation:

$$\mathbf{h}'_j = \sum_{i\in\mathcal{N}(j)} \boldsymbol{\alpha}_{ij} \odot \mathbf{q}_i\,, \tag{6.11}$$

The second GAT layer uses the output $\mathbf{h}'_j$ from the first layer as input. The attention mechanism is identical to the first layer, but includes an additional node update function that compresses the 512-dimensional representation to a scalar via an MLP followed by sigmoid activation:

$$\mathbf{h}''_j = \sigma\left(\texttt{node\_update}\left(\sum_{i\in\mathcal{N}(j)} \boldsymbol{\alpha}_{ij} \odot \mathbf{h}'_i\right)\right) \in [0,1]\,, \tag{6.12}$$

where σ is the sigmoid function. Table 6.1 summarizes the implementation details of both GAT layers.

After the second GAT block, the output graph retains the same topology as the input graph, but each node feature has dimension one and represents the predicted probability of the hit being signal (1) or background (0).

Stage	Description
Input graph	Nodes: (y, z) coordinates of hits Edges: fully connected graph with self-edges
GAT layer 1	`attention_query`: MLP([4, 8, 16, 32, 64, 128, 256, 512]) `attention_logit`: MLP([1024, 512] $\rightarrow$ [512]) + LeakyReLU
GAT layer 2	`attention_query`: Identity `attention_logit`: MLP([1024, 512] $\rightarrow$ [512]) + LeakyReLU `node_update`: MLP([512, 256, 128, 64, 32, 16, 8, 4, 2, 1]) + Sigmoid

Table 6.1: Details of the implementation of the GAT layers used for the end-to-end architecture for the task of muon tracking.

6.7 Differentiable Clustering Module

The output graph of the classification network is then combined with the original input graph and passed to a differentiable clustering module, which is defined as part of the same computational graph. In this step, hits are clustered layer by layer using a weighted mean, where the weights are given by the predicted signal probabilities. This mechanism naturally suppresses the contribution of background hits to the cluster position.

For each detector layer j, a cluster position $\bar{y}_j$ is computed as the weighted average of the hit positions.

$$\bar{y}_j = \frac{\sum_{i \in j} w_i \, y_i}{\sum_{i \in j} w_i} \,, \tag{6.13}$$

with w_i predicted label by the two-blocks GAT for each node i in the $j-$layer. The associated uncertainty $\sigma_{\bar{y}_j}$ is defined as the weighted root mean square (RMS) of the clustered hits. In addition, each cluster is assigned a weight $\mathbf{W}_j$, defined as the sum of the predicted signal probabilities of the hits contributing to the cluster. This choice ensures that clusters composed of many signal-like hits receive higher importance, while clusters dominated by background hits are assigned a lower weight.

In the special case where only a single hit is present in a layer, the cluster position is set equal to the hit position, the associated uncertainty is taken from the strip pitch resolution, and the cluster weight is given directly by the predicted signal probability of that hit.

6.8 Differentiable Circular Fitting

A second differentiable block performs the fitting of the clustered hits with a circular trajectory. The fit is based on a weighted version of the Kåsa method [80]. Layer-wise weights, defined as $\mathbf{W}_j/\sigma^2_{\bar{y}_j}$, are incorporated in each matrix element, thereby incorporating information about hit quality and signal purity directly into the fitting procedure.

The Kåsa method provides an analytic least-squares solution to the problem of fitting a circle to a set of points in the transverse plane. Considering a set of hit coordinates (x_i, y_i), a circular trajectory with center (a, b) and radius r satisfies the equation

$$(x - a)^2 + (y - b)^2 = r^2. \tag{6.14}$$

Expanding this expression gives

$$x^2 + y^2 - 2ax - 2by + a^2 + b^2 - r^2 = 0, \tag{6.15}$$

and introducing the algebraic parameters

$$D = -2a, \quad E = -2b, \quad F = a^2 + b^2 - r^2 \tag{6.16}$$

the circle equation can be rewritten in linear form as

$$x^2 + y^2 + Dx + Ey + F = 0. \tag{6.17}$$

For each hit (x_i, y_i) this yields the linear relation

$$Dx_i + Ey_i + F = -(x_i^2 + y_i^2). \tag{6.18}$$

So, stacking all N hits gives the matrix system

$$\mathbf{Ap} = \mathbf{b} \tag{6.19}$$

with

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & 1 \end{bmatrix}, \qquad \mathbf{p} = \begin{bmatrix} D \\ E \\ F \end{bmatrix}, \qquad \mathbf{b} = -\begin{bmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ \vdots \\ x_N^2 + y_N^2 \end{bmatrix}. \tag{6.20}$$

The least-squares solution is then the one which minimizes the sum of the squared residuals

$$S(\mathbf{p}) = \sum_{i=1}^{N} (b_i - (\mathbf{Ap})_i)^2 = (\mathbf{b} - \mathbf{Ap})^T (\mathbf{b} - \mathbf{Ap}). \tag{6.21}$$

Differentiating with respect to $\mathbf{p}$ and setting the derivative to zero give so the standard normal equations:

$$\mathbf{A}^T\mathbf{Ap} = \mathbf{A}^T\mathbf{b}, \tag{6.22}$$

whose solution is

$$\mathbf{p} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \tag{6.23}$$

To account for the varying quality of the hits, a weight w_i is introduced for each hit and for each event they are collected into a diagonal weight matrix

$$\mathbf{W} = \mathrm{diag}(w_1, w_2, .., w_N). \tag{6.24}$$

Once weights are included, the least-squares problem then minimizes the weighted residuals

$$S(\mathbf{p}) = \sum_{i=1}^{N} w_i \, (b_i - (\mathbf{Ap})_i)^2 = (\mathbf{b} - \mathbf{Ap})^T \mathbf{W} (\mathbf{b} - \mathbf{Ap}). \tag{6.25}$$

Following the same procedure as before, differentiating $S(\mathbf{p})$ with respect to $\mathbf{p}$ and setting the derivative to zero gives the weighted normal equations:

$$\mathbf{A}^T\mathbf{WAp} = \mathbf{A}^T\mathbf{Wb}. \tag{6.26}$$

Solving for $\mathbf{p}$ yields the weighted solution:

$$\mathbf{p} = (\mathbf{A}^T\mathbf{WA})^{-1}\mathbf{A}^T\mathbf{Wb}. \tag{6.27}$$

In this formulation, hits with larger weights contribute more strongly to the fit, effectively pulling the fitted circle toward the higher-quality measurements, while hits with smaller weights have a reduced influence. Is also important to notice that with this formulation the circle fitting problem is reduced to a fully analytic matrix operation, requiring only standard linear algebra operations like matrix multiplication, transposition, and inversion, allowing the solution to be computed in an efficient and differentiable way for any set of input hits and weights. As a result, gradients can be propagated through the fitting step during training.

Once the algebraic coefficients $\mathbf{p} = [D, E, F]^T$ are obtained, the geometric circle parameters are recovered using the Kåsa relations:

$$a = -\frac{D}{2}, \quad b = -\frac{E}{2}, \quad R = \sqrt{a^2 + b^2 - F}. \tag{6.28}$$

These parameters can then be directly used to compute the reconstructed transverse muon momentum $\hat{p}_\phi$ through the relation $\hat{p}_\phi = 0.3\, B\, r$, with B magnetic field strength and r the fitted radius.

6.9 Model Outputs and Loss function

For each input event, the model produces two outputs:

- node-level predictions corresponding to the signal/background classification of detector hits;
- an event-level prediction of the transverse momentum reconstructed from the fitted trajectory.

Both outputs are incorporated into a composite loss function, defined as follows:

$$\hat{\phi} = \arg\min_{\phi} \mathcal{L}(\hat{p}_\phi, w_{i,\phi}) = \tag{6.29}$$

$$= \arg\min_{\phi} \sum_{\text{events}} \left[\alpha \frac{(\hat{p}_\phi - p)^2}{p^2} + \frac{1}{N_{\text{hits}}} \sum_{i=1}^{N_{\text{hits}}} \text{BCE}(w_{i,\phi}, y_{i,\text{truth}}) \right]. \tag{6.30}$$

Here, the momentum p in the loss term reflects the true p_T value for each event, while $y_{i,\text{truth}}$ designates the truth labels corresponding to each individual node within the graph.

Together, these components define a fully differentiable, end-to-end tracking model in which all stages of the reconstruction pipeline are optimized simultaneously with respect to the final physics task.

The hyperparameter α balances the regression and classification components and it is empirically set to 0.5. During the early training epochs, α is set to 0 so that only the binary cross-entropy (BCE) term is initially considered. This two-stage approach was observed to yield a more rapid convergence of the training process.

Figure 6.3 provides a schematic overview of the full model, visually summarizing the architecture described in the previous sections and presented in [75]. Each input event is represented as a fully connected graph, with node features corresponding to the hit coordinates (y_i, z_i). This graph is processed by the Graph Attention Network (GAT) f_ϕ, producing an output graph with the same topology, where the node features correspond to the predicted signal probabilities $w_{i,\phi} = f_\phi(\mathbf{x}_i)$.

The predicted node labels are then combined with the input features and passed through the differentiable clustering module, yielding layer-wise clustered hit positions $\bar{y}$, their associated uncertainties $\sigma_{\bar{y}}$, and cluster weights W. These clusters are subsequently fitted with a circular trajectory using the Kåsa method, and the resulting radius is used to compute the reconstructed transverse momentum $\hat{p}_{T,\phi}$.

As all components of the architecture described in the previous section are defined in a fully differentiable manner, gradients from the final loss function — which compares both the node-level predictions and the reconstructed $\hat{p}_{T,\phi}$ with the corresponding truth information encoded in the associated truth graph — can be back-propagated through the entire model.

This enables the parameters ϕ of the full architecture described previously, which is inspired by Graph Attention Networks but constitutes a standalone model, to be optimized end-to-end with respect to the complete tracking task.

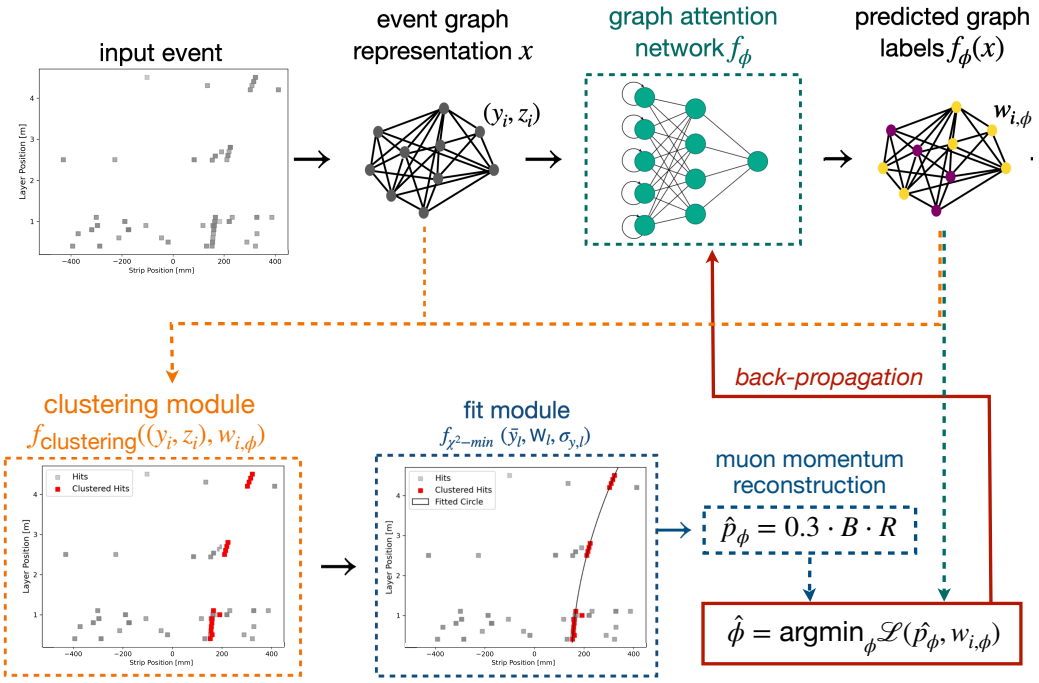


Figure 6.3: Schematic representation of the end-to-end model developed for track reconstruction. Both the clustering and fitting routines depend on the GAT learnable parameters through the predicted node labels $w_{i,\phi}$. The regression of the particle's p_T is back-propagated to the patter recognition module as visualized in the figure and described in more details within the text [75].

6.10 Training setup

The model was trained on a dataset of 10^4 events and evaluated on an independent test sample of 9×10^3 events. Optimization during training was performed using the Adam optimizer [81], together with an `OPTAX` learning rate scheduler that gradually decays the learning rate.

As discussed previously, the training procedure is divided into two phases. In the first phase, only the classification loss is minimized, corresponding to setting the parameter $\alpha = 0$. This choice is motivated by training stability considerations: the reconstruction of the transverse momentum depends strongly on the classification performance, and including the regression loss from the beginning was observed to slow down convergence. Once the classification loss has sufficiently stabilized, the regression term is activated by setting $\alpha = 0.5$, allowing the model to optimize both classification and physics-informed regression simultaneously.

Figure 6.4 shows the evolution of the training and validation loss terms as a function of the epoch, separately for the hit classification task and the transverse momentum regression. Solid lines correspond to the training sample, while dashed lines indicate the evaluation on the validation sample. The vertical dashed line marks the epoch at which the classification-only training was considered stable and no longer significantly optimizable. At this point, the transverse momentum regression term was activated by setting $\alpha = 0.5$ in the composite loss function.

During the initial training phase, only the classification loss is explicitly minimized. Nevertheless, the p_T loss values shown in this regime are obtained by passing, at each epoch, the optimized output graphs through the differentiable clustering and fitting modules for both the training and validation datasets. Although no direct optimization is performed on the p_T term in this phase, an improvement in the regression loss is naturally observed as a consequence of the improved hit classification, which leads to more accurate cluster reconstruction and a more

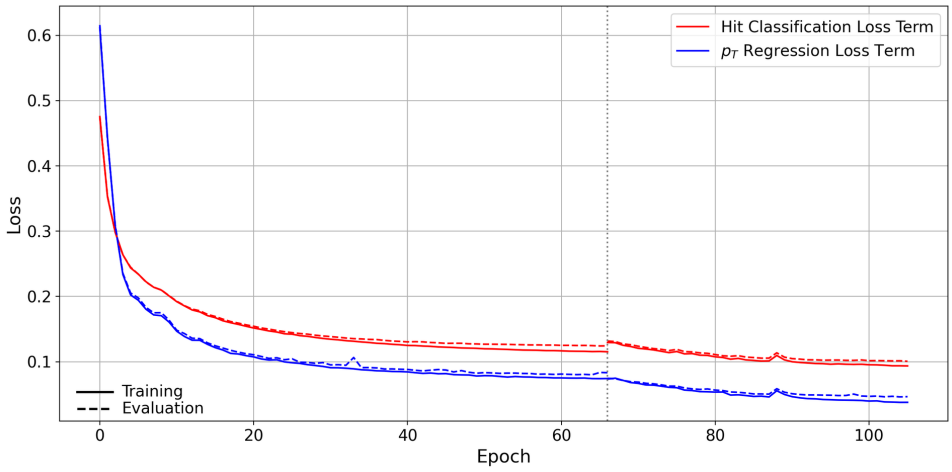


Figure 6.4: Training and validation loss trends for the hits classification and p_T prediction separately. The vertical dashed line defines the two different regimes, before it only the classification task is optimized ($\alpha = 0$), then also the p_T loss term is activated and weighted with $\alpha = 0.5$.

stable circular fit.

Once the p_T regression term is included in the loss function, a further decrease is observed not only in the regression loss itself but also in the classification loss. This behavior highlights the benefit of the end-to-end differentiable formulation: even though the classification and regression terms are not directly coupled, they share the same set of trainable parameters through the differentiable clustering and fitting blocks. As a result, optimizing the global objective leads to a more favorable minimum for both tasks compared to optimizing the classification loss alone.

6.11 Results

The performance of the proposed approach is assessed using two complementary observables: the hit-classification accuracy and the resolution of the reconstructed transverse momentum p_T. To quantify the impact of the end-to-end differentiable training, the results are compared to those obtained with a reference model that shares the same network architecture but is trained exclusively for the node-level classification task.

This baseline model is selected from the final training epochs before the parameter α is switched from 0 to 0.5. As such, it reflects a more traditional, factorized reconstruction strategy, in which hit classification is optimized independently, and the subsequent clustering and fitting steps are applied without being explicitly included in the training objective. Importantly, the same clustering and fitting modules are used for both models, ensuring that any observed differences in performance can be attributed to the training strategy rather than to differences in the downstream reconstruction algorithms.

In the following discussion and figures, the model trained with the full end-to-end objective is referred to as the end-to-end model, while the classification-only reference is denoted as the baseline model.

6.11.1 Hit Classification Accuracy

For the node-classification task, performance is evaluated using output score histograms for signal and background hits, and by constructing the ROC curve obtained by scanning different

score thresholds, which represents the background rejection as a function of the signal efficiency. ROC distributions are generated for both the baseline and the end-to-end model, and reported in Figure 6.5, showing that the physics information inclusion within the model explicitly through the p_T loss term results also in improved classification performance thanks to the differentiability of the implementation.

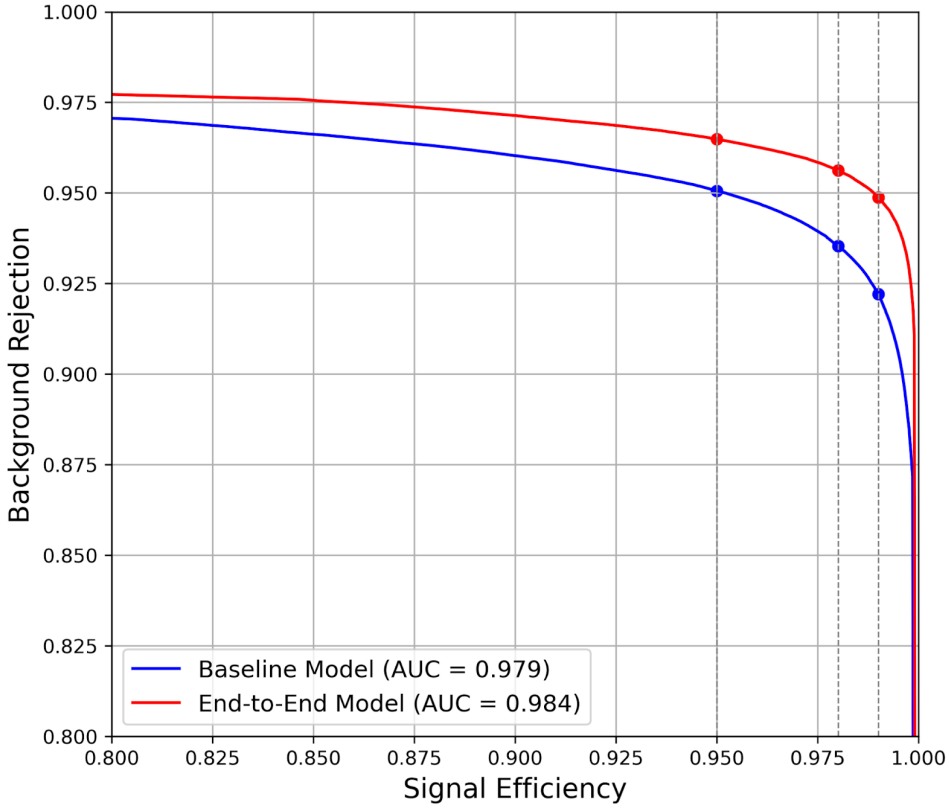


Figure 6.5: ROC distirbutions for the hit classification task for both the baseline and the end-to-end models for tracking. The curve is obtained scanning different cuts on the predicted signal and background distributions and measuring the amount of selected signal hits (with the score higher than the cut value) and the discarded background ones (score lower than the cut). End-to-end model higher background rejections clearly shows how the differentiable inclusion of physics prior boosts the accuracy of the hit classification task. The three defined operating points, corresponding to signal efficiencies of 95%, 98%, and 99%, corresponds to a background rejection consistently improved by 2-3% for the end-to-end model with respect to the baseline one [75].

A complementary study was carried out to quantify the spatial resolution of the reconstructed cluster positions. This analysis focuses on the residuals between the reconstructed cluster coordinates and the true muon impact point on each detector layer, providing a direct measure of the clustering accuracy.

The residual distributions were evaluated for three different configurations: an idealized reference obtained using truth-weighted clustering, the baseline model trained exclusively on the hit-classification task, and the proposed end-to-end model. The truth-weighted configuration represents an upper-performance bound, in which only signal hits are used with ideal weights.

To define a reference resolution scale for the detector, the residual distribution obtained with truth-weighted clustering was fitted with a Gaussian function, and the width of this distribution was used to identify the detector's core resolution region. For each of the three configurations, the fraction of reconstructed clusters whose residuals fall within this core region was then computed.

Compared to the baseline model, the end-to-end approach exhibits a significant improvement in spatial reconstruction performance. In particular, the fraction of clusters reconstructed within the core resolution window increases by 12.7% when using the end-to-end model. This result demonstrates that the joint optimization of hit classification, clustering, and fitting, enabled by the fully differentiable formulation, leads to more accurate cluster position estimates, even in the absence of direct supervision on cluster-level observables.

6.11.2 Reconstructed p_T Resolution

The impact of the end-to-end training strategy on momentum reconstruction is evaluated by comparing the accuracy of the reconstructed transverse momentum p_T across the same three models used in the hit classification studies. Reconstructed p_T values are compared between the end-to-end model outputs and those obtained by using either the truth information for each node or the baseline model predictions as inputs to the clustering and fitting modules. This setup ensures that the comparison isolates the effect of the training strategy and information flow, rather than differences in the reconstruction algorithms themselves.

The performance of the p_T reconstruction is assessed through the residual distribution of q/p_T, where q denotes the muon charge. This choice provides a linear sensitivity to the track curvature in the magnetic field, making it a natural observable for evaluating the quality of the circular fit. For each reconstruction strategy, the q/p_T residuals are computed by comparing the reconstructed values to the corresponding truth p_T values.

To robustly quantify the resolution, the width σ^* of each residual distribution is computed following the procedure described in Ref. [82]. The σ^* is defined as half the width of the smallest interval, centered on the median of the distribution, containing 68% of the events. This estimator is less sensitive to non-Gaussian tails than a standard RMS, providing a stable measure of the core momentum resolution.

The resulting residual distributions and their corresponding σ^* values, shown in Figure 6.6, enable a direct comparison of the three approaches, highlighting the performance gains achieved by the end-to-end model relative to both the baseline and the idealized reference cases.

A more detailed assessment of momentum reconstruction is performed by analyzing the p_T residuals in bins of true transverse momentum. The study uses 5 GeV-wide bins in the 20–50 GeV range, which is particularly relevant for the momentum spectrum of interest in the considered detector configuration. The truth p_T of each muon serves as the reference for computing residuals in each bin.

For every p_T bin, the resolution is quantified using the σ^* quantity for each of the three configuration, and the trend is compared for each case. This differential figure of merit is particularly informative, as it probes the momentum-dependent behavior of the reconstruction, revealing potential non-uniformities or biases that might be obscured in an inclusive resolution measurement. As shown in Figure 6.7, the end-to-end model exhibits systematically smaller σ^* values across all bins, closely approaching the performance obtained using truth-based inputs. This demonstrates the improved resolution achieved by incorporating the differentiable p_T objective in the training.

6.12 Comments and Conclusions

This chapter presented a fully differentiable, end-to-end approach to charged-particle tracking, in which hit classification, clustering, and momentum reconstruction are jointly optimized within a unified framework. The main motivation behind this work was to explore whether modern machine learning techniques, particularly graph neural networks combined with differentiable programming, can provide performance gains in tracking similar to those achieved in

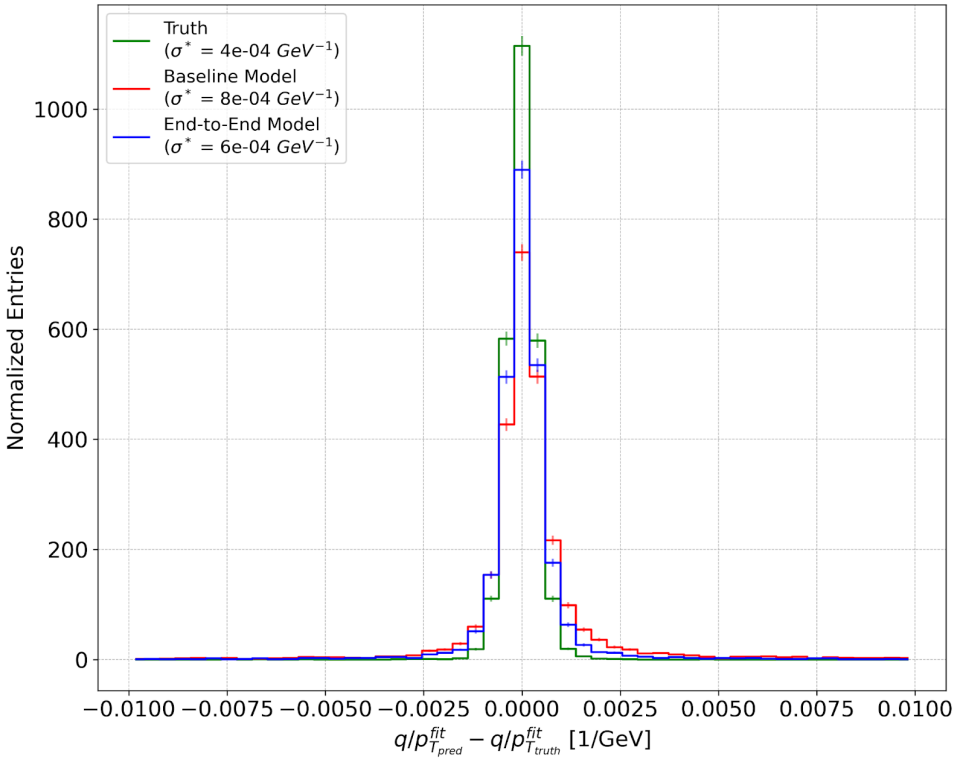


Figure 6.6: Normalized q/p_T residual distributions, with q being the charge of the reconstructed muon. The three distributions correspond to using the truth labels as input weight in the clustering step, using the baseline model prediction and those from the proposed end-to-end model. The end-to-end model demonstrates an improved accuracy in p_T regression, approaching the limit of an ideal detector [75].

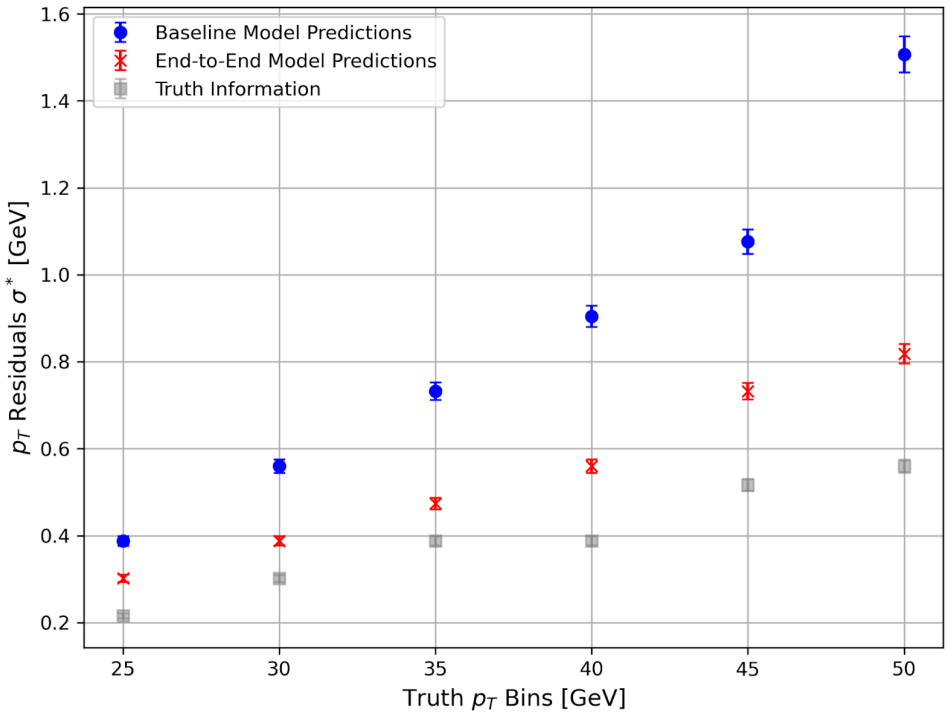


Figure 6.7: p_T residual distribution as a function of the truth p_T values and defined through the σ^* quantity [75].

jet tagging. The proposed model leverages Graph Attention Networks to encode correlations between detector hits, a differentiable clustering module to combine hits into layer-wise clusters, and a differentiable circular fitting procedure to reconstruct the transverse momentum p_T. By defining all components within a single differentiable computational graph, gradients from the final physics objective can propagate throughout the entire pipeline, allowing the model to simultaneously optimize for both hit classification and momentum estimation. End-to-end differentiable training enhances both hit classification and p_T reconstruction. By jointly optimizing these tasks through a shared set of parameters, the model achieves better performance in both areas compared to a traditional factorized approach. Including physics-informed priors in the loss function, such as transverse momentum regression, further improves performance by allowing the network to exploit correlations between hit patterns and track curvature, leading to more precise cluster positions and momentum estimates.

The differentiable design of the clustering and fitting modules ensures effective information flow: feedback from the final momentum reconstruction propagates to upstream layers, guiding the classification network to produce more informative outputs. Additionally, the improvements are consistent across the relevant momentum spectrum, as differential resolution studies show that gains are not confined to specific p_T ranges.

Overall, this work demonstrates the feasibility and advantages of fully differentiable, end-to-end tracking models. The results indicate that such models can approach the performance of an ideal detector, improving hit classification, clustering, and momentum reconstruction compared to conventional factorized methods. While the current study focuses on a simplified single-muon scenario, the implemented strategy could be applied to more complex detector geometries, offering a promising path for future developments in both offline reconstruction and online triggering.

Conclusion

In the era of precision measurements of the Higgs boson, an accurate understanding of its dominant decay modes is essential to test the Standard Model and probe possible deviations. This thesis focuses on the most probable decay channel, $H \to b\bar{b}$, which, despite its large branching ratio, remains experimentally challenging due to the overwhelming QCD background. In this context, the physics of b-jets using the full ATLAS Run-2 dataset was studied through two main aspects: the calibration of the boosted GN2X tagger, crucial for reliable heavy-flavour identification at high transverse momentum, and the investigation of the VH Higgs production mode at high momenta when using the calibrated tagger.

The calibration program was developed along two complementary directions. The first approach relied exclusively on Monte Carlo simulations through the so-called adjusted MC method. I completely developed, within the already existing ATLAS flavour tagging frameworks, the methodology for extracting the correction factors to be applied on MC samples in order to align the tagger performance with respect to the one observed on data. As the GN2X tagger classification relies on variables related to the tracks within the large radius jet used for reconstructing the entire Higgs decay, a set of variations related to track quantities was included in MC samples in order to quantify the tagger performance variation with each of them. Since the set of variations affecting the tracking inputs to the tagger was limited, and several of the dominant systematic uncertainties relevant for flavour tagging were not available, the resulting calibration wasn't considered fully reliable, but marks an important step forward towards future developments.

To complement this strategy, a second calibration approach based on $Z \to b\bar{b}$+jets events was developed using the full Run 2 dataset. In this case, I worked on the scale factors extraction through a comparison of the signal yields measured in data and simulation. The yields were obtained via a signal-plus-background fit on the leading large-R jet mass distributions following a dedicated background modelling procedure which introduced additional systematic uncertainties related to the choice of templates used to model the QCD background in the fit.

Within this preliminary setup, the final scale factors were derived as a function of the transverse momentum of the leading large-R jet and subsequently included in the $VH, H \to b\bar{b}$ boosted analysis, obtaining values between 1 and 1.15 with associated uncertainties of $\sim$10%. The inclusion of the GN2X tagger in the $VH, H \to b\bar{b}$ boosted analysis, exploiting its strong discrimination power, made it possible to extend the boosted regime towards lower transverse momenta. In particular, kinematic regions that were treated as resolved in the latest public ATLAS results can now be incorporated into the boosted analysis, thereby enlarging the accessible phase space. The analysis strategy was closely aligned with that of the most recent ATLAS publication, relying on a multivariate approach in which boosted decision trees (BDTs) are trained to discriminate signal from background separately in each leptonic channel, and their output distributions are subsequently used as inputs to the statistical inference.

Within this framework, the adoption of the GN2X tagger leads to a significant improvement in sensitivity compared to previous results based on the DL1r tagger. In particular, an increase of 23% is observed when using BDT output scores, resulting in an inclusive expected sensitivity of 7.67 across all leptonic channels and the extended p_T range. Motivated by this improvement

at the event selection level, additional studies were performed to assess the potential benefit of incorporating GN2X-related information directly into the BDT inputs. Different practical implementations were explored, all indicating a further gain in expected sensitivity with respect to the baseline approach. These results highlight the potential of a tighter integration of advanced flavour-tagging information within multivariate analyses and open promising directions for future optimizations.

The final part of the thesis presented an independent research project exploring a differentiable, end-to-end machine-learning approach to charged-particle track reconstruction. The proposed framework defines each component of the reconstruction procedure in a differentiable way, allowing the entire pipeline to be trained jointly. In this setting, physics priors, such as the presence of a circular trajectory describing the motion of a charged particle in a magnetic field, can be directly incorporated into the learning objective. The results demonstrate that including such physically motivated constraints leads to improved performance in both hit classification and transverse-momentum regression when compared to a more traditional factorized approach. This study therefore serves as a proof of concept for the potential of differentiable programming techniques and physics prior inclusion in the development of future track reconstruction algorithms.

Appendix

Spurious signal test results

This appendix collects the full set of results obtained from the spurious signal tests. The material is organised in bins of transverse momentum, with separate sections corresponding to the 450–500 GeV, 500–600 GeV, and 600–1000 GeV intervals. For each p_T bin, four tables are presented, corresponding to different background modelling functions: a third-degree polynomial (Poly3), and exponential-polynomial functions of second, third, and fourth degree (Expoly2, Expoly3, and Expoly4, respectively).

Within each configuration, the results are further detailed for all considered working points, defined in terms of their associated mistag efficiencies. For each case, the mean (μ) and width (σ) are extracted by fitting with a Gaussian the distributions of the μ/σ values obtained from the spurious signal tests performed on toy datasets. In addition, the mean and standard deviation of the μ-only distributions are reported, together with the percentage of failed fits.

The choice of the background model is performed independently for each working point and p_T bin by selecting the configuration for which the mean of the fitted μ/σ distribution is below 0.2; in cases where no configuration satisfies this requirement, a looser threshold of 0.35 is applied. Configurations fulfilling the 0.2 (0.35) criterion are highlighted in green (yellow). Among the selected candidates, the final background model used in the signal-plus-background fit is chosen as the one with the smallest number of free parameters.

For the chosen configuration, the fitted mean value of the μ-only distribution is taken as the corresponding contribution to the scale factor systematic uncertainty associated with the choice of the background modelling function. Final selection is reported in Table 4.7.

p_T in [450, 500] GeV

Table 6.2: poly3 for p_T in [450, 500] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	0.4	1.09	0.04	0.11	1
0.30	0.37	1.03	0.04	0.1	0
0.37	0.74	0.94	0.08	0.1	0
0.46	0.27	0.92	0.03	0.1	0
0.58	0.47	0.93	0.05	0.11	0
0.74	0.72	1.03	0.09	0.13	0
0.94	0.59	1.04	0.08	0.14	0
1.25	0.77	0.86	0.11	0.12	0
1.55	0.79	0.89	0.12	0.14	1

Table 6.3: expoly2 for p_T in [450, 500] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.34	1.08	-0.03	0.1	0
0.30	-0.25	1.0	-0.02	0.1	0
0.37	-0.24	0.96	-0.02	0.09	1
0.46	-0.61	0.91	-0.06	0.09	0
0.58	-0.72	0.93	-0.08	0.1	0
0.74	-0.5	1.04	-0.06	0.12	0
0.94	-0.62	1.03	-0.07	0.13	0
1.25	-0.27	0.92	-0.04	0.12	0
1.55	-0.58	0.96	-0.08	0.14	1

Table 6.4: expoly3 for p_T in [450, 500] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	0.13	1.07	0.01	0.1	0
0.30	0.07	1.02	0.01	0.1	1
0.37	0.3	0.97	0.03	0.1	0
0.46	-0.15	0.96	-0.02	0.1	0
0.58	-0.08	0.92	-0.01	0.1	0
0.74	0.13	1.05	0.01	0.12	1
0.94	-0.06	1.03	-0.01	0.13	0
1.25	0.15	0.89	0.02	0.12	0
1.55	0.02	0.91	0.0	0.13	0

Table 6.5: expoly4 for p_T in [450, 500] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.0	1.0	0.0	0.11	0
0.30	0.0	1.08	0.0	0.13	1
0.37	0.25	1.0	0.03	0.13	0
0.46	-0.17	0.91	-0.02	0.12	1
0.58	-0.01	1.0	-0.0	0.13	1
0.74	0.09	1.07	0.02	0.15	1
0.94	0.03	1.03	0.0	0.15	0
1.25	0.08	0.86	0.01	0.14	0
1.55	-0.07	0.87	-0.01	0.16	0

p_T in [500, 600] GeV

Table 6.6: poly3 for p_T in [500, 600] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.11	0.89	-0.01	0.09	1
0.30	-0.07	1.09	-0.01	0.11	0
0.37	0.03	0.91	0.0	0.1	0
0.46	0.03	1.12	0.01	0.12	1
0.58	-0.17	1.07	-0.02	0.12	0
0.74	-0.57	0.92	-0.07	0.11	0
0.94	-0.6	0.96	-0.08	0.13	1
1.25	-0.57	1.09	-0.09	0.16	1
1.55	-0.7	1.01	-0.11	0.15	1

Table 6.7: expoly2 for p_T in [500, 600] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	0.36	0.94	0.03	0.09	1
0.30	0.31	1.07	0.03	0.1	0
0.37	0.16	0.96	0.02	0.09	0
0.46	0.09	1.14	0.01	0.11	1
0.58	-0.02	1.08	-0.0	0.12	0
0.74	-0.44	0.94	-0.05	0.11	1
0.94	-0.64	0.94	-0.08	0.12	1
1.25	-0.43	1.03	-0.06	0.14	0
1.55	-0.4	0.94	-0.06	0.13	2

Table 6.8: expoly3 for p_T in [500, 600] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.03	0.89	-0.0	0.09	0
0.30	0.04	1.09	0.0	0.1	0
0.37	0.03	0.94	0.0	0.09	0
0.46	-0.03	1.14	-0.0	0.12	1
0.58	-0.23	1.03	-0.03	0.11	0
0.74	-0.68	0.94	-0.08	0.11	0
0.94	-0.76	0.93	-0.09	0.12	1
1.25	-0.62	1.04	-0.09	0.14	0
1.55	-0.72	0.96	-0.1	0.14	1

Table 6.9: expoly4 for p_T in [500, 600] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.04	0.86	-0.0	0.1	1
0.30	-0.04	1.09	-0.01	0.13	0
0.37	0.11	0.95	0.01	0.11	0
0.46	0.04	1.05	0.01	0.13	1
0.58	-0.07	0.98	-0.01	0.13	0
0.74	-0.25	0.96	-0.04	0.13	0
0.94	-0.23	1.09	-0.04	0.16	1
1.25	-0.18	1.0	-0.04	0.17	0
1.55	-0.19	1.09	-0.04	0.2	1

p_T in [600, 1000] GeV

Table 6.10: poly3 for p_T in [600, 1000] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.15	0.97	-0.01	0.09	0
0.30	0.02	0.88	0.0	0.08	0
0.37	-0.07	0.96	-0.01	0.09	0
0.46	-0.18	1.0	-0.02	0.1	1
0.58	-0.16	0.97	-0.02	0.1	0
0.74	-0.61	0.86	-0.07	0.1	0
0.94	-0.41	0.93	-0.05	0.12	1
1.25	-0.61	0.99	-0.08	0.13	0
1.55	-0.9	1.12	-0.13	0.16	0

Table 6.11: expoly2 for p_T in [600, 1000] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.07	0.99	-0.01	0.08	4
0.30	0.31	0.94	0.03	0.08	5
0.37	0.14	0.95	0.01	0.09	5
0.46	0.01	1.09	0.0	0.1	6
0.58	-0.03	0.91	-0.0	0.09	4
0.74	-0.63	0.90	-0.07	0.1	0
0.94	-0.24	0.96	-0.03	0.11	0
1.25	-0.33	1.02	-0.04	0.13	4
1.55	-0.22	1.07	-0.03	0.14	3

Table 6.12: expoly3 for p_T in [600, 1000] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.13	0.96	-0.01	0.09	0
0.30	0.09	0.9	0.01	0.08	0
0.37	-0.01	0.96	-0.0	0.09	0
0.46	-0.17	1.04	-0.02	0.1	0
0.58	-0.14	0.93	-0.02	0.1	0
0.74	-0.79	0.89	-0.08	0.1	0
0.94	-0.5	0.95	-0.06	0.11	0
1.25	-0.7	1.0	-0.09	0.13	0
1.55	-0.85	1.1	-0.12	0.15	0

Table 6.13: expoly4 for p_T in [600, 1000] GeV

QCD Eff [%]	μ fit (μ/σ)	σ fit (μ/σ)	μ fit (μ)	σ fit (μ)	Failed Fit [%]
0.25	-0.12	0.9	-0.02	0.11	0
0.30	0.05	0.99	0.01	0.12	0
0.37	-0.06	0.97	-0.01	0.12	0
0.46	-0.07	0.97	-0.01	0.13	0
0.58	-0.07	0.87	-0.01	0.12	0
0.74	-0.06	0.92	-0.01	0.13	1
0.94	-0.04	1.01	-0.0	0.16	0
1.25	-0.03	0.98	-0.01	0.17	0
1.55	-0.03	1.06	-0.01	0.2	0

References

[1] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1), 2012.

[2] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1), 2012.

[3] CERN. CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, 2017.

[4] S. Dawson et al. Higgs physics: It ain't over till it is over. *Physics Reports*, 816, 2019.

[5] L.R. Evans. The Large Hadron Collider Project. 1997.

[6] E. Mobs. The CERN accelerator complex . 2018. `https://cds.cern.ch/record/2636343`.

[7] ATLAS Collaboration. ATLAS data quality operations and performance for 2015-2018 data-taking. *JINST*, 15(04), 2020.

[8] I. Neutelings. TikZ examples by Izaak Neutelings, 2021. `https://tikz.net/author/izaak/`.

[9] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3, 2008.

[10] ATLAS Collaboration. Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment. `https://cds.cern.ch/record/2723878`, 2020.

[11] M.Capeans et al. ATLAS Insertable B-Layer Technical Design Report. Technical report, 2010.

[12] ATLAS Collaboration. ATLAS Approved Plots: Trigger and Data Acquisition, 2025. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ`.

[13] ATLAS Collaboration. Athena: Software framework for the ATLAS experiment at CERN, 2019. `https://gitlab.cern.ch/atlas/athena`.

[14] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *NIM-A*, 262(2), 1987.

[15] ATLAS Collaboration. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *EPJC*, 77(5), 2017.

[16] Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2015.

[17] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$TeV. *EPJC*, 79(8), 2019.

[18] M. Cacciari et al. The anti-kt jet clustering algorithm. *JHEP*, 04, 2008.

[19] M. Cacciari et al. FastJet user manual. *EPJC*, 72(3), 2012.

[20] R. Bouquet. *Physics with b-jets using the ATLAS Run 2 data: from calibration to Higgs boson couplings measurements in the $VH, H \to b\bar{b}$ channel.* PhD thesis, 2023.

[21] ATLAS Collaboration. Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS. 2020.

[22] L. Santi. *The Beauty and Charm Yukawa Couplings of the Higgs Boson with the ATLAS Detector at the LHC - 4D Tracking, Particle Flow and Jet Flavour Reconstruction Algorithms Development.* PhD thesis, University of Rome La Sapienza, 2024.

[23] ATLAS Collaboration. Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *JHEP*, 2013(9), 2013.

[24] D. Krohn et al. Jet trimming. *JHEP*, 2010(2), 2010.

[25] Particle Data Group. Review of particle physics. *Prog. Theor. Exp. Phys.*, 2024, 2024.

[26] ATLAS Collaboration. Secondary vertex finding for jet flavour identification with the ATLAS

detector. Technical report, CERN, Geneva, 2017.

[27] G. Piacquadio et al. A new inclusive secondary vertex algorithm for b-jet tagging. *JPCS*, 119(3), 2008.

[28] ATLAS Collaboration. Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. 2017.

[29] ATLAS Collaboration. Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment. 2017.

[30] ATLAS Collaboration. Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS. 2020.

[31] ATLAS Collaboration. Transforming jet flavour tagging at ATLAS. 2025. Submitted to: Nature Communications.

[32] JM. Butterworth et al. Jet substructure as a new Higgs search channel at the LHC. 2008.

[33] S. Marzani et al. *Looking Inside Jets: An Introduction to Jet Substructure and Boosted-object Phenomenology*. Springer International Publishing, 2019.

[34] ATLAS Collaboration. Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle x in hadronic final states using $\sqrt{s} = 13$TeV pp collisions with the ATLAS detector. *Phys. Rev. D*, 108(5), 2023.

[35] ATLAS Collaboration. Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS. 2020.

[36] ATLAS Collaboration. Measurement of the b-jet identification efficiency with the $p_{\mathrm{T}}^{\mathrm{rel}}$ method in multi-jet events using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. 2022.

[37] M. Cacciari et al. The catchment area of jets. *JHEP*, 2008(04), 2008.

[38] ATLAS Collaboration. Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS. 2023.

[39] ATLAS Collaboration. Calibration of light-flavour b-jet mistagging rates using ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV. 2018.

[40] ATLAS Collaboration. Alignment of the ATLAS Inner Detector in Run-2. *Eur. Phys. J. C*, 80(12), 2020.

[41] Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2015.

[42] Zach Marshall. Re-re-defining the Standard QCD Di-Jet Samples. 2015.

[43] ATLAS Flavour Tagging group. Training Dataset Dumper, 2024. `https://gitlab.cern.ch/aft/algorithms/training-dataset-dumper`.

[44] ATLAS Collaboration. Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2021.

[45] ATLAS Collaboration. Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using 139fb^{-1} of ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV. *EPJC*, 83(8), 2023.

[46] T. Gleisberg et al. Event generation with Sherpa 1.1. *JHEP*, 02, 2009.

[47] E. Bothmann et al. Event generation with Sherpa 2.2. *SciPost Phys.*, 7, 2019.

[48] J. Alwall et al. The automated computation of tree-level and NLO differential cross sections, and their matching to parton shower simulations. *JHEP*, 07, 2014.

[49] T. Sjostrand et al. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191, 2015.

[50] R. Frederix and S. Frixione. Merging meets matching in MC@NLO. *JHEP*, 12, 2012.

[51] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *Journal of Instrumentation*, 14(12), 2019.

[52] ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. *EPJC*, 81(7), 2021.

[53] J. Wang. An intuitive tutorial to gaussian process regression. *IEEE Computer Society*, 25(4), 2023.

[54] P. Berta. ATLAS jet and missing-ET reconstruction, calibration, and performance. Technical

report, CERN, Geneva, 2016.

[55] ATLAS Collaboration. Cross-section measurements of the Higgs boson decaying into a pair of tau-leptons in proton-proton collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Rev. D*, 99(7), 2019.

[56] CMS Collaboration. Observation of the Higgs boson decay to a pair of leptons with the CMS detector. *Physics Letters B*, 779, 2018.

[57] ATLAS Collaboration. Evidence for the $H \to b\bar{b}$ decay with the ATLAS detector. *JHEP*, 2017(12), 2017.

[58] CMS Collaboration. Observation of Higgs Boson Decay to Bottom Quarks. *PRL*, 121(12), 2018.

[59] ATLAS Collaboration. Measurement of the Higgs boson production in association with top quarks in multilepton final states in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, 2025.

[60] A.K. Nayak et al. Searches for ttH production at CMS, 2018.

[61] LHC Working Group. Report from Working Group 2. *CERN Yellow Rep. Monogr.*, 7, 2019.

[62] ATLAS Collaboration. Measurements of WH and ZH production with Higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in 13 TeV pp collisions with the ATLAS detector. *JHEP*, 2025(4), 2024.

[63] ATLAS Collaboration. ATLAS flavour-tagging algorithms for the LHC Run 2 *pp* collision dataset. *Eur. Phys. J. C*, 83, 2023.

[64] G. Callea. Updates on 2L MET trigger study. `https://indico.cern.ch/event/1035473/contributions/4358873/attachments/2242627/3802690/LegacyMeeting_trigger2L_1105.pdf`.

[65] ATLAS Collaboration. Object definition of $VH(bb/cc)$ Legacy Analysis. Technical report, CERN, 2022. `https://cds.cern.ch/record/2843084`.

[66] Y. Coadou. *Boosted Decision Trees*. WORLD SCIENTIFIC, 2022.

[67] Berger et al. Simplified template cross sections – Stage 1.1 and 1.2. *SciPost Physics Community Reports*, 2026.

[68] ATLAS Collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(11), 2019.

[69] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13TeV. *EPJC*, 76(5), 2016.

[70] Caillou et al. ATLAS ITk Track Reconstruction with a GNN-based pipeline. Technical report, CERN, Geneva, 2022.

[71] P. de Castro et al. INFERNO: Inference-aware neural optimisation. *CPCs*, 244, 2019.

[72] N. Simpson et al. neos: End-to-end-optimised summary statistics for high energy physics. *JPCS*, 2438(1), 2023.

[73] R.E.C. Smith et al. Differentiable vertex fitting for jet flavour tagging, 2023.

[74] ATLAS Collaboration. Optimizations of the ATLAS ITk GNN reconstruction pipeline. Technical report, CERN, Geneva, 2025.

[75] L. Rambelli S. Rosati A. Coccaro, F. Armando Di Bello and C. Schiavi. Learning to reconstruct: A differentiable approach to muon tracking at the LHC, 2025. `https://arxiv.org/abs/2512.01647`.

[76] Benoit Lefebvre. Muon Spectrometer Phase-I Upgrade for the ATLAS Experiment: the New Small Wheels project. Technical report, CERN, Geneva, 2018.

[77] JAX developers. *JAX Documentation*. Google, 2025. `https://docs.jax.dev/en/latest/`.

[78] DeepMind. The DeepMind JAX Ecosystem, 2020.

[79] P. Veličković et al. Graph attention networks, 2018.

[80] A. Al-Sharadqah and N.I. Chernov. Error analysis for circle fitting algorithms. *EJS*, 3, 2009.

[81] P. Diederik et al. Adam: A method for stochastic optimization, 2017. `https://arxiv.org/abs/1412.6980`.

[82] ATLAS Collaboration. Studies of the muon momentum calibration and performance of the ATLAS detector with pp collisions at $\sqrt{s} = 13$TeV. *EPJC*, 83(8), 2023.