



DEPARTMENT OF PHYSICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis

**A new algorithm for the identification
of boosted $Z \rightarrow e^+e^-$ decays
for heavy resonance searches
with the ATLAS detector at the LHC**

Florian Johannes Kiwit



MAX-PLANCK-GESELLSCHAFT





DEPARTMENT OF PHYSICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis

**A new algorithm for the identification
of boosted $Z \rightarrow e^+e^-$ decays
for heavy resonance searches
with the ATLAS detector at the LHC**

**Ein neuer Algorithmus zur Identifizierung
von geboosteten $Z \rightarrow e^+e^-$ Zerfällen
für Suchen nach schweren Resonanzen
mit dem ATLAS Detektor am LHC**

Author:	Florian Johannes Kiwit
Supervisor:	Dr. Dominik Duda
Advisor:	Prof. Hubert Kroha
Submission Date:	November 2022

I confirm that this master's thesis is my own work and I have documented all sources and material used.

Munich, November 2022

Florian Johannes Kiwit

To my family

Abstract

The identification of W , Z , and Higgs bosons with large transverse momenta is crucial in many searches for new heavy resonances. Thus far, the development of algorithms for the tagging of boosted bosons focuses on the reconstruction and identification of hadronic boson decays, while no dedicated algorithm to identify boosted $Z \rightarrow e^+e^-$ decays exists. The performance of the standard electron reconstruction and identification algorithms degrades with increasing transverse momentum of the e^+e^- pairs.

This thesis describes the development and performance of a new approach for reconstructing and identifying highly boosted $Z \rightarrow e^+e^-$ decays with the ATLAS detector. A $Z \rightarrow e^+e^-$ candidate decay is reconstructed via a single jet, clustered via the anti- k_t algorithm using a radius parameter of 0.4. For the identification of the $Z \rightarrow e^+e^-$ decay, a deep neural network, trained on the jet properties including inner detector and calorimeter information, is used. Signal efficiencies and background rejection rates are evaluated as a function of the jet kinematics. Furthermore, the mass and transverse momentum response scales and resolutions are studied.

Finally, the $Z \rightarrow e^+e^-$ identification and reconstruction approach is tested in the search for a Z' boson based on Monte Carlo simulations of the data taken with the ATLAS detector during the LHC Run 2. Expected exclusion limits on the production cross section times branching ratio at 95% confidence level are determined.

Contents

Acknowledgments	iii
Abstract	iv
1. Introduction	1
2. The Standard Model of Particle Physics	4
2.1. The Particle Content of the Standard Model	4
2.2. The Standard Model Lagrangian	6
2.3. Beyond the Standard Model of Particle Physics	12
2.3.1. Model independent Search Strategies	13
3. Experimental Setup	16
3.1. The Large Hadron Collider	16
3.2. The ATLAS Detector	18
3.2.1. Coordinate System	18
3.2.2. The Inner Detector	19
3.2.3. The Calorimeter System	21
3.2.4. The Muon System	23
3.2.5. The Trigger and Data Acquisition System	24
3.3. Reconstruction and Identification of Physic Objects	24
3.3.1. Electrons	24
3.3.2. Muons	29
3.3.3. Jet Clustering	29
3.3.4. Jet Energy Correction	35
3.3.5. Missing Transverse Energy	38
4. Introduction to Deep Learning	39
4.1. The Artificial Neuron	40
4.2. Architectures	41
4.3. Training	43
4.4. Regularisation	46

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays	48
5.1. Shortcomings of the Standard $Z \rightarrow e^+e^-$ Identification and Reconstruction	49
5.2. The Monte Carlo Dataset	53
5.3. Object Reconstruction	54
5.4. Input Variables	67
5.5. Network Architecture and Training Procedure	73
5.6. Performance	77
5.7. Variable Importance	87
5.8. Comparison to the Standard Method	89
6. The Search for new heavy Resonances	90
6.1. Monte Carlo Samples	91
6.2. Analysis Strategy and Event Selection	93
6.3. Statistical Analysis	96
6.4. Upper Limits on the Cross Section times Branching Ratio	98
7. Concluding Remarks	100
A. Appendix	102
A.1. Track Reconstruction Studies	102
A.2. Jet Resolution Studies	103
A.3. Planar Flow Derivation	109
A.4. Performance on the Training Set	110
List of Figures	112
List of Tables	121
Bibliography	122

1. Introduction

The Standard Model of Particle Physics (SM) explains the nature of matter and fundamental forces more precisely than any other theory by describing its fundamental constituents as quantum fields. Its predictions were confirmed in multiple experiments advancing to the smallest reachable distances and the highest reachable energies. After searching for decades for the particles predicted by the SM [1–6], the last puzzle piece was found with the discovery of the Higgs boson in 2012 [7, 8].

Despite all its achievements, the SM fails to describe some phenomena observed in nature: three of the four fundamental interactions are included in the SM, the electromagnetic, the weak and the strong interaction. However, the fourth interaction, gravity, has yet to be described via a quantum field theory. Furthermore, astrophysical observations strongly suggest that the particle content of the SM accounts only for about 5% of the universe; the nature of the so far unobserved content, i.e. *dark matter* and *dark energy*, still needs to be understood. Since the corresponding particles do not interact electromagnetically, they are considered *dark*. Another limitation of the SM is the electroweak hierarchy problem, which describes that the mass of the Higgs boson is much smaller than expected.

There are a variety of theories attempting to explain these open questions by introducing new physics. Many extensions of the SM [9–13] agree on the assumption that new heavy vector bosons W' and Z' must exist that predominantly decay into pairs of bosons (WW , ZZ , WZ , Wh , Zh). These new particles are expected to have similar properties as the mediators of the weak interaction.

The search for new heavy particles is carried out at particle accelerators such as the Large Hadron Collider (LHC) [14] at CERN near Geneva. With multi-purpose detectors such as ATLAS [15] and CMS [16], particle collisions with unrivalled energies are recorded. Perhaps, particles such as the Z' have already been produced, and it is just a matter of isolating their signal from the backgrounds.

Evolving accelerator and detector technologies allow us to study increasingly complex detector signatures in growing data samples; however, to exploit the full potential of the gathered data, software advancements are needed as well. *Deep learning* is a powerful tool that has experienced rapid growth in the last decade and promises to improve

particle research in a wide range of fields. Applications include classification tasks, anomaly detection and event generation.

Searches for new heavy vector bosons have always been a core component of the ATLAS [15] and CMS [16] physics programme, and by now, these searches probe for signs of new physics at a mass scale extending to several TeV [17–24]. At such high energies, the final state particles are strongly collimated, and dedicated reconstruction and identification algorithms are required. Thus far, the development of boosted boson tagging algorithms has focused on the reconstruction and identification of hadronic boson decays [25–29] or decays into boosted $\tau^+\tau^-$ systems [30], while no dedicated algorithm to identify boosted $Z \rightarrow e^+e^-$ decays exists.

The ATLAS standard techniques for electron reconstruction [31] are based on matching the track of a charged particle to energy clusters in the calorimeter system, while electron identification is based on a likelihood procedure that combines information on electromagnetic shower shapes, track quality, and track–cluster matching in order to separate prompt electrons from relevant backgrounds. Both, reconstruction and identification procedures are highly-performant over a large kinematic range. However, the electron reconstruction and identification performance in highly boosted $Z \rightarrow e^+e^-$ decays is significantly limited. In fact, several recent diboson resonance searches based on channels targeting leptonic Z boson decays were found to be limited by the inefficiency to reconstruct and identify highly boosted $Z \rightarrow e^+e^-$ decays [19–21]. The degradation effects for large Z boson momenta (i.e. for strongly collimated e^+e^- systems) are driven by, e.g. the merging of energy clusters and distortion of shower shapes and track quality observables.

This thesis describes a novel approach for reconstructing and identifying boosted $Z \rightarrow e^+e^-$ decays using a new algorithm based on a deep neural network. $Z \rightarrow e^+e^-$ decays are reconstructed as a particle jet and identified based on the jet properties, including information from the tracking detector and the calorimeter system of the ATLAS detector. The documentation of the development and performance of this new algorithm dedicated to reconstructing and identifying highly boosted $Z \rightarrow e^+e^-$ decays has been published by the ATLAS collaboration [32]. The algorithm is implemented into a search for new heavy vector bosons, and the results are interpreted within the predictions of the Heavy Vector Triplet model [33].

The thesis is structured as follows: Chapter 2 introduces the SM, and Chapter 3 summarises the experimental setup at ATLAS. Further, an introduction to Deep Learning is detailed in Chapter 4. Then, the limitations of the standard methods and the deployment of the novel approach for probing boosted $Z \rightarrow e^+e^-$ decays are detailed in Chapter 5. Next, the application of this new algorithm in the search for Vh resonances

is described in Chapter 6. Finally, Chapter 7 provides a discussion, an outlook and some concluding remarks.

2. The Standard Model of Particle Physics

The Standard Model of particle physics (SM) describes the foundations of modern particle physics in the language of quantum field theory. The SM postulates subatomic particles (see Figure 2.1), which are assumed to be fundamental, i.e., they have no substructure. The fundamental constituents are described by fields, and the excitations of these fields are the particles that will be introduced in the following. In 2012 the last puzzle piece was found with the discovery of the Higgs boson [7, 8].

Natural Units Natural units are physical units based only on universal physical constants. The SI units are then expressed in terms of physical constants, like the reduced Planck constant $\hbar$ and the speed of light c . These constants are then set to 1, leading to simplified equations of motion. Consequently, energy, momentum, and mass have the same unit (eV). In this thesis, natural units are used. Further, according to the Einstein summation convention, an index variable appearing twice implies a summation.

2.1. The Particle Content of the Standard Model

The elementary particles are unique in intrinsic properties, such as mass, electric charge, spin, and colour charge. Depending on the spin quantum number, particles are separated into fermions and bosons. Fermions are associated with matter, while bosons are the force carriers. In addition to gravitation, three known elementary interactions exist in nature: the electromagnetic, weak and strong interaction. Each is governed by the exchange of vector bosons.

Bosons Photons are the carriers of electromagnetic force; in everyday life, photons are for example encountered as light or electromagnetic waves for wireless communication. The Z and W bosons are the mediators of the weak interaction most prominent in explaining the radioactive beta decay. Furthermore, the weak interaction governs hydrogen fusion into helium, powering the sun's thermonuclear process. The strong

interaction binds atomic nuclei and also their subatomic constituents. The field excitations of the strong interaction are the gluons. As the electromagnetic interaction occurs between particles that carry electric charge by exchanging photons, the strong interaction occurs between particles carrying colour charge by exchanging gluons. Besides the vector bosons, the SM particle content includes the Higgs boson. The Higgs mechanism was independently introduced by Englert and Brout [35] as well as by Higgs [36] and explains how massive particles acquire their mass by interaction with the Higgs field. [37]

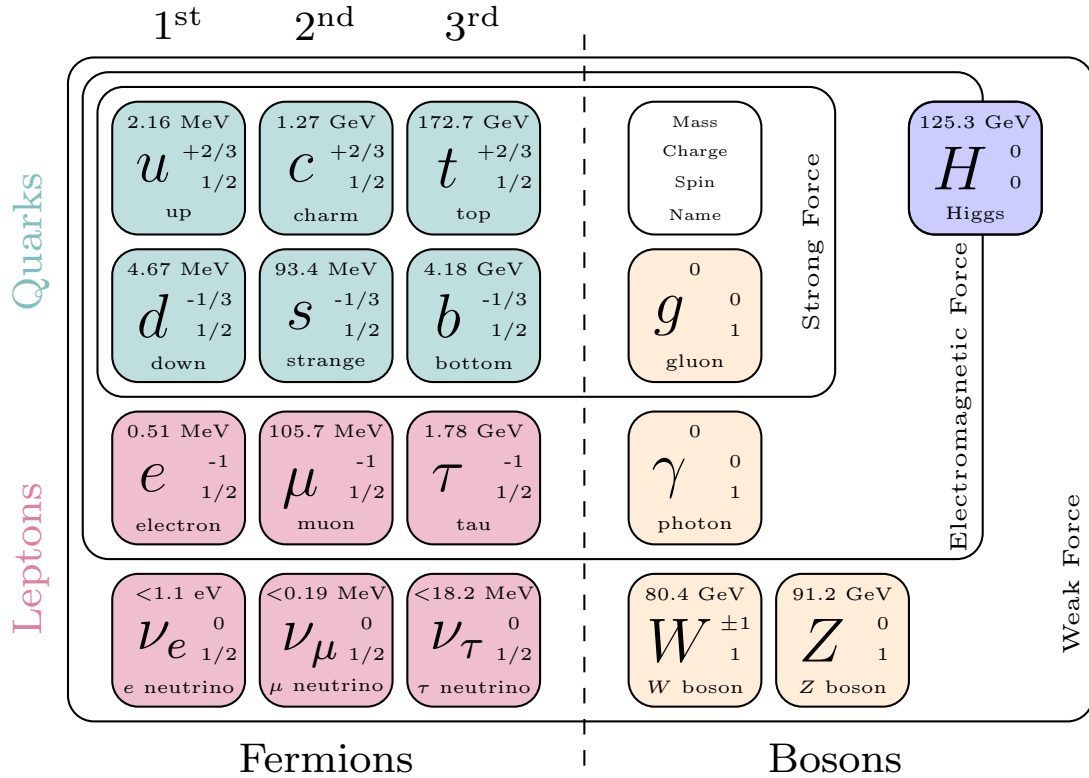


Figure 2.1.: Particle content of the Standard Model of particle physics. The two types of fermions, quarks (leptons) in green (red), are ordered in the three generations. The gauge bosons are in red, and the Higgs boson in yellow. The physical constants are published by the Particle Data Group [34].

Fermions All fermions undergo the weak force and have a half-odd-integer spin; depending on whether they have a colour charge, they are classified as quarks or leptons.

As quarks carry colour charge, they couple to gluons. They exist in six different quark flavours; two quarks are assigned to each of the three generations. All atoms in the periodic table are compound systems of particles of the first generation: the protons and neutrons are the building blocks of the atomic nucleus and consist of up and down quarks. The second and third generations contain particles differing only by mass in comparison to the particles of the first generation. The up, charm, and top quarks are up-type quarks with an electric charge of $+2/3 e$; the down, strange, and bottom quarks are down-type quarks with an electric charge of $-1/3 e$. Due to their electric charge, all quarks interact electromagnetically, and their masses increase from the first to the third generation. Each quark exists in each colour, leading to a total number of 18 quarks.

Fermions not interacting via the strong force form the class of leptons. They are significantly lighter than quarks and are also assigned to the three generations. The first generation includes the electron with an electric charge $-1 e$ and the electron neutrino with an electric charge $0 e$. The second generation is formed by the muon and muon neutrino, and the third by the tau and tau neutrino.

Additionally, every elementary particle with a non-zero charge quantum number (such as electric charge) is associated with an antiparticle. The antiparticles resemble the particles in mass and spin but carry opposite charges. [37]

2.2. The Standard Model Lagrangian

In the SM, the *Lagrangian density*¹ $\mathcal{L}$ describes kinematics, potentials, and interactions of fundamental particles. The SM is a local gauge theory: by requiring invariance of the Lagrangian under transformations of the $SU(3) \times SU(2) \times U(1)$ symmetry group, the gauge fields are introduced. The excitations of the gauge fields are the carriers of the electromagnetic, weak and strong interactions. To derive the equations of motion, the principle of stationary action, the so-called *Hamilton's principle*, is applied to the Lagrangian. It is a central axiom in modern physics, claiming that physical systems behave so that the action integral

$$S = \int d^4x \mathcal{L}$$

¹In the following, the Lagrangian density will be simply referred to as Lagrangian.

is stationary and is then translated into the Euler-Lagrange equations. By applying the Euler-Lagrange equation

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} = 0,$$

the model's Lagrangian is translated into the equations of motion. In the following, the structure of the SM Lagrangian is motivated. According to the Lagrangian, as proposed by Paul Dirac, fermions are massive spin-1/2 particles.

$$\mathcal{L}_{Dirac} = i\bar{\psi}(\gamma^\mu \partial_\mu - m)\psi$$

The local phase transformation U(1) is defined as $\psi(x) \rightarrow \psi'(x) = e^{iq\chi(x)}\psi(x)$. As shown in the following, $\mathcal{L}_{Dirac}$ is not symmetric under the transformation.

$$\begin{aligned} \mathcal{L}_{Dirac} \rightarrow \mathcal{L}'_{Dirac} &= ie^{-iq\chi}\bar{\psi}(\gamma^\mu \partial_\mu - m)e^{iq\chi}\psi \\ &= \mathcal{L}_{Dirac} + ie^{-iq\chi}\bar{\psi}\gamma^\mu(\partial_\mu e^{iq\chi})\psi \\ &= \mathcal{L}_{Dirac} - q\bar{\psi}\gamma^\mu(\partial_\mu \chi)\psi \end{aligned}$$

To establish local U(1) symmetry, the normal derivative is substituted by the covariant derivative $\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu$ with well defined gauge transformation properties of the electromagnetic four-potential $A_\mu \rightarrow A'_\mu(x) = A_\mu - \partial_\mu \chi$. Replacing the normal derivative with the covariant and adding the kinetic term of the electromagnetic field leads to the Lagrangian of quantum electrodynamics which describes a Dirac fermion, a massless spin-1 particle (photon), and their interaction.

$$\mathcal{L}_{QED} = i\bar{\psi}(\gamma^\mu \partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + q\bar{\psi}\gamma^\mu\psi A_\mu$$

The kinetic term $F_{\mu\nu}F^{\mu\nu}$ depends on the electromagnetic field tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ and is invariant under U(1) transformations. Adding a mass term for the four-potential would lead to a problem since it is not invariant under U(1) transformations. However, this problem needs no further attention since photons are massless.

$$\frac{1}{2}m^2 A_\mu A^\mu \rightarrow \frac{1}{2}m^2 (A_\mu - \partial_\mu \chi)(A^\mu - \partial^\mu \chi) \neq \frac{1}{2}m^2 A_\mu A^\mu$$

However, following the same strategy for the W and Z boson causes a problem due to their experimentally proven mass. The remedy lies in the introduction of the Higgs mechanism, which spontaneously breaks the electroweak $SU(2)_L \times U(1)_Y$ symmetry. Particles acquire their mass by interacting with the Higgs field. The Higgs field needs four degrees of freedom to provide the longitudinal degrees of freedom to gauge bosons of the weak interaction and a massive scalar boson, i.e. the Higgs boson. The four degrees of freedom are arranged in a doublet, each component consisting of a complex scalar field. Since the Higgs mechanism must generate the masses of the charged $W^\pm$ boson and the neutral Z boson, one scalar field needs to be neutral, while the other scalar field needs to be charged. This leads to the Higgs field

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}.$$

Then, the Lagrangian of the Higgs field is defined via:

$$\mathcal{L}_{Higgs} = (\partial_\mu \phi)^\dagger (\partial_\mu \phi) - V(\phi), \quad (2.1)$$

with the potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2.$$

The parameter λ must be positive to guarantee a finite minimum of the potential. However, there is no such restriction for the parameter μ^2 . But when choosing the parameter $\mu^2 < 0$, the Higgs potential has a unique shape as shown in Figure 2.2. Due to its unique shape, the Higgs potential is also referred to as *Mexican Hat Potential*. The Higgs potential has degenerated minima for field configurations fulfilling the relation

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}.$$

By selecting a value that minimises the Higgs potential, the Higgs mechanism *spon-*

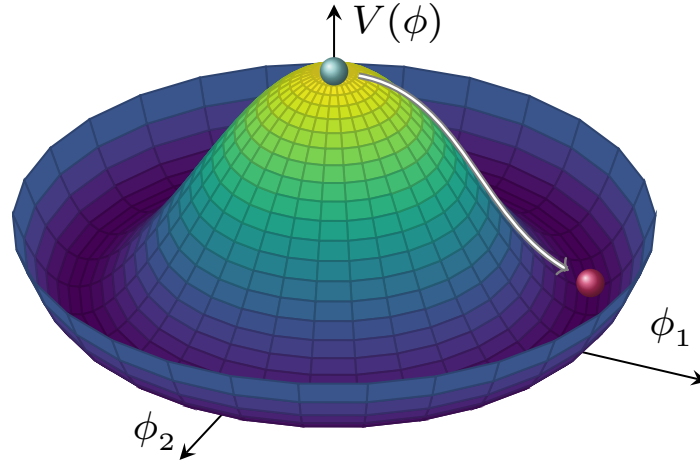


Figure 2.2.: Illustration of the Higgs potential. The SM Higgs potential is not two, but four-dimensional. Selecting a value of the degenerated minimum at the bottom of the potential spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry.

taneously breaks the electroweak symmetry. This point corresponds to the physical vacuum state. The photon needs to be massless after symmetry breaking; hence the charged component of the Higgs field has a vacuum expectation value (vev) of 0.

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$$

The fields are expanded about the vacuum expectation value.

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + h + i\phi_4 \end{pmatrix} \quad (2.2)$$

Substituting Equation 2.2 into Equation 2.1 leads to the massive physical Higgs field h , and the three non-physical mass-less Goldstone bosons. The gauge freedom of the Lagrangian allows transferring the non-physical Goldstone bosons to the longitudinal freedom of the gauge bosons of the weak interaction. After the so-called *Uniform Gauge*,

the expansion of the Higgs field about the vev is given by

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (2.3)$$

The corresponding Lagrangian is the Glashow-Weinberg-Salam (GWS) model [38–40]. The derivative is substituted by the covariant derivative

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_W \mathbf{T} \cdot \mathbf{W}_\mu + ig' \frac{Y}{2} B_\mu,$$

with the generators of the SU(2) group $\mathbf{T} = \frac{1}{2}\boldsymbol{\sigma}$, the three gauge fields of the weak interaction $\mathbf{W}_\mu$, the weak hypercharge Y , the gauge field B_μ of the U(1) symmetry, and the coupling constants g_W and g' . The lower component of the Higgs doublet is neutral and has weak isospin $I_W^{(3)} = \frac{1}{2}$. Hence, the weak hypercharge of the GWS model is $Y = 2(Q - I_W^{(3)}) = 1$.

By inserting the expansion about the vev and the covariant derivative into the kinetic term of the Lagrangian, the mass terms of the gauge fields (i.e. the quadratic terms) are accessible. After transforming the equation, the mass terms of the gauge boson fields are given by

$$\frac{1}{8}v^2 g_W^2 \left(W_\mu^{(1)} W^{(1)\mu} + W_\mu^{(2)} W^{(2)\mu} \right) + \frac{1}{8}v^2 \left(g_W W_\mu^{(3)} - g' B_\mu \right) \left(g_W W^{(3)\mu} - g' B^\mu \right). \quad (2.4)$$

In the Lagrangian, the masses of the $W^{(1)}$ and $W^{(2)}$ fields appear as $\frac{1}{2}m_W^2 W_\mu^{(1)} W^{(1)\mu}$ and $\frac{1}{2}m_W^2 W_\mu^{(2)} W^{(2)\mu}$. Therefore, the mass of the $W^\pm$ boson can be expressed solely by the vev and the electroweak coupling constant $m_W = \frac{1}{2}g_W v$.

The mass terms of the physical Z boson and the photon are still hidden but can be accessed by some basic linear algebra. The second addend from Equation 2.4 can be transformed to

$$\frac{v^2}{8} \left(g_W W_\mu^{(3)} - g' B_\mu \right) \left(g_W W^{(3)\mu} - g' B^\mu \right) = \frac{v^2}{8} \begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} \underbrace{\begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix}}_{\mathbf{M}} \begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix}.$$

By comparing the diagonalised mass matrix $\mathbf{M}$

$$\frac{v^2}{8} \begin{pmatrix} A_\mu & Z_\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & g_W^2 + g'^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix},$$

with the mass matrix as expected in the Lagrangian, the mass of the photon $m_A = 0$ and the Z boson $m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2}$ can be identified. The corresponding eigenvalues are

$$A_\mu = \frac{g'W_\mu^{(3)} + g_W B_\mu}{\sqrt{g_W^2 + g'^2}}, \text{ and } Z_\mu = \frac{g_W W_\mu^{(3)} - g' B_\mu}{\sqrt{g_W^2 + g'^2}},$$

which corresponds to the physical photon field A_μ and the neutral field of weak interaction Z_μ respectively. Besides vector bosons, the Higgs boson couples to fermions, which is described by the Yukawa term. The Yukawa terms for each fermion (except neutrinos) are given by

$$\mathcal{L} = -m_f \bar{f}f - \frac{m_f}{v} \bar{f}fh,$$

with $m_f = \frac{g_f}{\sqrt{2}}$. Hence, the couplings of the fermion fields f to the Higgs field h are proportional to the fermion's mass.

Additionally, the SM Lagrangian describes the strong interaction of quantum chromodynamics (QCD). In QCD, quarks are described by triplets with colour charge red (R), green (G), and blue (B). Additionally, each colour charge exists as anti-colour ($\bar{R}$, $\bar{G}$, $\bar{B}$). QCD is evolved by requiring symmetry of the SM Lagrangian under a SU(3) symmetry group. The eight Gell-Mann matrices are one possible set of generators of the SU(3) symmetry group. Consequently, there are eight additional gauge fields, the so-called gluon octet. Gluons carry a combination of colour and anti-colour and can therefore couple to themselves. One crucial aspect of QCD is *asymptotic freedom*: as the energy density increases, the strength of the strong interaction decreases. Thus, particles with colour charge can exist as free particles at high-enough energies or low-enough distances only. At small energies, however, quarks and gluons do not occur isolated but in colourless compound states, the compound states are referred to as *confinement*. Colourless states are built by combining all three colours or the combination of colour and anti-colour.

Further, the SM includes quark flavour mixing as described by the Cabibbo–Kobayashi–Maskawa (CKM) matrix and neutrino oscillations as described by the Pontecorvo–

Maki–Nakagawa–Sakata matrix (PMNS matrix). [41–46]

2.3. Beyond the Standard Model of Particle Physics

Many questions on the behaviour of nature remain unclear from the particle physics point of view. Prominent examples are dark matter, the naturalness problem, or the description of gravitation as a quantum field theory. [41]

Dark Matter Astrophysical observations strongly suggest that the universe contains not only the matter described by the SM but also *dark matter*, which does not interact electromagnetically but gravitationally. The galaxy rotation curve describes the dependence of the orbital velocity of stars on the distance from the galaxy centre. According to the Doppler shift of spectral lines in the stellar spectra of the Milky Way, the rotational velocity differs from the predicted value by gravitational theory applied to the present matter.

Gravitation As dark matter, gravity is not described within the SM framework. At energies reached by modern particle colliders such as the LHC, the influence of gravitation is negligible, and the SM’s predictions remain highly accurate. For very high energies, however, the gravitational force becomes more relevant since the coupling is proportional to the mass.

Higgs-hierarchy problem In the SM, the bare mass of the Higgs boson $m_{H,\text{bare}}$ undergoes corrections from loop interactions with other particles. The observed Higgs mass (see Figure 2.1) is given by the sum

$$m_{H,\text{observed}}^2 = m_{H,\text{bare}}^2 + \Delta m_H^2,$$

with the quantum loop correction at leading order $\Delta m_H^2 = \frac{1}{16\pi^2}(-6g_t^2 + g^2 + \lambda^2) \cdot \Lambda^2$ (g_t , g , and λ are the t -quark Yukawa coupling, the couplings of weak gauge bosons, and the Higgs self-coupling respectively). The cut-off value Λ can be considered as the energy scale to which the SM maintains validity. The difference of several magnitudes between the Higgs mass and the Planck scale leads to the necessity that the coupling parameters need to be determined very precisely ($\mathcal{O}(m_H^2/m_p) = 10^{-32}$). The *fine-tuning* of the bare Higgs mass is considered *unnatural* and known as the *Higgs-hierarchy problem*. It motivates the possibility of new physics at the TeV scale, such as new particles, compositeness, or extra dimensions.

2.3.1. Model independent Search Strategies

As detailed, the SM is not a full description of nature, and physics beyond the SM must exist. Various theories are attempting to explain these new phenomena. However, when probing the predictions of these theories in data, some complications arise:

1. The theories depend on free parameters that need to be tuned by experiments. These parameters must be set to a fixed value to test the theories. The scan over the parameter space scales exponentially and is computationally not feasible.
2. Various theories explain the same effects, and often there is no evidence of which theory is true.
3. The theories are incomplete because the broad assumptions about new physics can not be translated into a concrete model.

A potential remedy lies in a model-independent strategy for resonance searches based on a simplified phenomenological Lagrangian, as proposed in [33], bridging the gap between theory and data (see Figure 2.3). Various explicit models are condensed into a simplified model with broad assumptions, but robust qualitative predictions like the existence of new particles. Further, resonance searches are not sensitive to the entire set of model parameters, which reduces the combinatorial large space of model parameters.

Heavy Vector Triplet Framework Various BSM theories like Technicolor [47–49], Composite Higgs [50, 51], and Little Higgs [52] extend the SM by introducing a new $SU(2)_L$ symmetry, including electroweak spin-1 particles, referred to as heavy vector triplet (HVT) framework, in the following. Technicolor and Composite Higgs agree on the assumption that the Higgs boson is not a fundamental particle: in technicolour, the

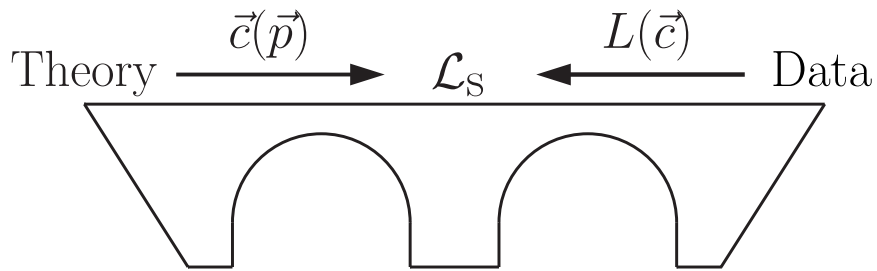


Figure 2.3.: Pictorial view of the Bridge Method. Theory and data meet at the central pillar. [33]

Higgs is a compound state of new fermions, the techniquarks; in Composite Higgs, the Higgs is a bound state of new strong interactions. In Little Higgs, the Higgs boson is a pseudo-Nambu-Goldstone boson from symmetry breaking at a TeV scale. In the HVT framework, the exotic electroweak spin-1 particles have large branching ratios to the Z, W and Higgs boson.

In addition to the SM fields and interactions, three new gauge fields V_μ^a with $a \in 0, 1, 2$ are introduced in the HVT framework. The excitations of these fields are one neutral heavy spin-one particle referred to as Z' and two neutral heavy spin-one particles referred to as $W^{\pm'}$. The fundamental particles' kinematics, potentials, and interactions are described by the phenomenological Lagrangian $\mathcal{L}_V$. The free parameters of $\mathcal{L}_V$ governing the coupling with the SM particles are of particular interest and will be commented on in the following. The parameter g_V describes the typical magnitude of the new heavy vector triplet interactions. The parameter c_H controls the V interactions with SM bosons, and c_F governs direct interactions with fermions.

For testing the theory, the free parameters need to be fixed to a value. Therefore, two benchmark models are defined: both have $c_F \sim 1$, but they differ by the value of c_H . For *Model A*, such that the similar branching ratios to fermions and bosons are predicted, while for *Model B*, c_H is set to a value such that the coupling to bosons is significantly larger than to fermions.

Figure 2.4 shows the branching Ratios for the two body decays of the neutral vector V^0 as well as the total width for the two Models in dependence of the resonance mass. [33, 53]

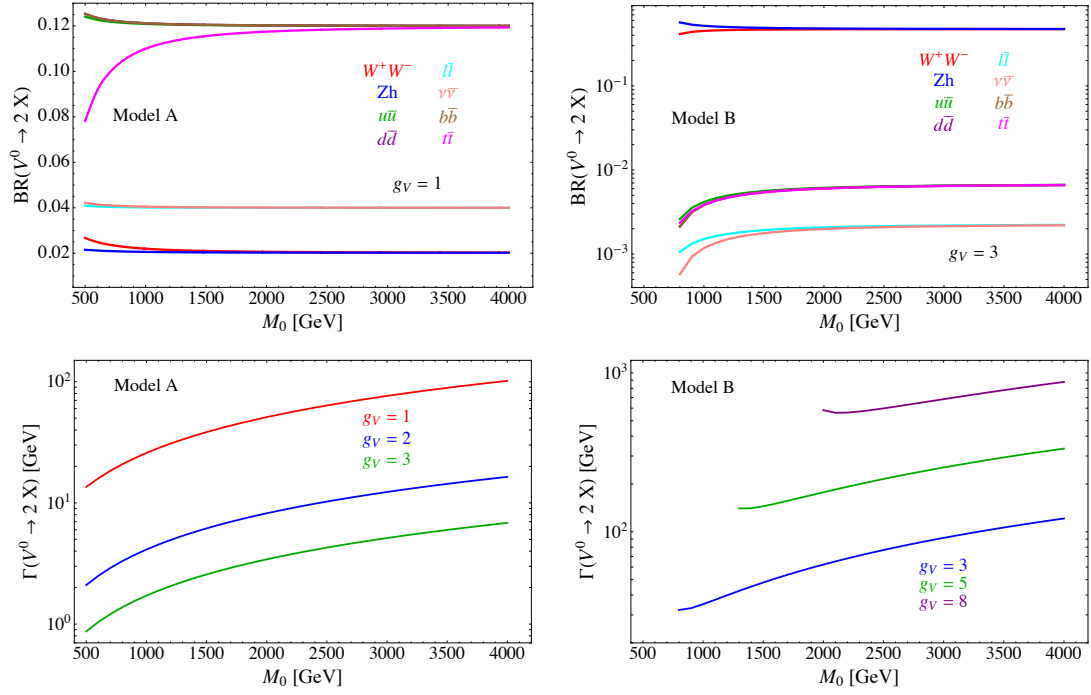


Figure 2.4.: Upper panel: Branching Ratios for the two body decays of the neutral vector V^0 for the benchmarks $A_{g_V=1}$ (left) and $B_{g_V=3}$ (right). Lower panel: Total widths corresponding to different values of the coupling g_V in the models A (left) and B (right). [33]

3. Experimental Setup

The analysis presented in this thesis is based on simulated data from ATLAS (A Toroidal LHC Apparatus) [15], a detector at the Large Hadron Collider (LHC) [14]. The LHC and ATLAS are operated by CERN (Conseil européen pour la recherche nucléaire), which is a major international research facility near Geneva in Switzerland, founded in 1954. With the LHC, hypotheses within the Standard Model framework are tested, and phenomena beyond the SM are searched. Alongside groundbreaking insights into the laws of nature such as the Higgs boson discovery in 2012 [7, 8], innovations like the World Wide Web were invented at CERN.

3.1. The Large Hadron Collider

In September 2008, the most powerful and largest man-made particle accelerator started operations. The LHC has a circumference of 27 km and is situated in the former Large Electron Positron (LEP) tunnel at the border of Switzerland and France. The quest for new particles starts rather unspectacular - with a small gas cylinder providing hydrogen. By ionising the hydrogen, protons are created and accelerated by passing a series of pre-accelerators. The acceleration in the LHC is realised with radiofrequency (RF) cavities operated at a frequency of 400 MHz. The 16 RF cavities are cooled to a temperature of 4.5 K to reach the superconducting state. The system forcing the particles on their circular trajectory is based on 1232 superconducting magnets. As the RF cavities, the magnets operate in the superconducting state, which is reached at a temperature of 1.9 K. The protons in the accelerated beams of proton bunches reach energies of up to 7 TeV, and the beam consists of about 2800 bunches with 10^{11} protons per bunch. At the intersection points of the two beams, the four detectors ALICE (A Large Ion Collider Experiment) [54], ATLAS, CMS (Compact Muon Solenoid) [16] and LHCb (Large Hadron Collider Beauty Experiment) [55] are placed (see Figure 3.1). ATLAS and CMS are multipurpose detectors with a similar physics programme aiming to measure physical observables with unreached precision and search new physics processes. Alice is specialised in investigating quark-gluon plasma, while LHCb focuses on CP-violating processes with b -mesons. Next to protons, the LHC can also be operated with fully

3. Experimental Setup

ionised lead atoms. In proton mode, the LHC is designed to reach center of mass energies of up to 14 TeV; so far, data has been gathered at energies up to 13.6 TeV.

Besides the center of mass energy, the luminosity provided by an accelerator is a crucial characteristic. It quantifies the process rate normalised to the probability of one specific process to occur and is referred to as cross section σ . The *Luminosity* can be represented solely by collider-specific parameters and is defined as

$$\mathcal{L} = \frac{dN}{dt} \frac{1}{\sigma} = \frac{n_b N_1 N_2 f \gamma}{4\pi \epsilon_n \beta^*} F, \quad (3.1)$$

with the number of bunches n_b , the number of particles per bunch in each of the two beams N_1 and N_2 , the revolution rate f , the Lorentz factor γ , the beam emittance ϵ taking the occupied area in phase space into account, the beta function β describing the transverse size of the beam and the geometric factor F depending on the beam crossing angle. The integrated luminosity $\mathcal{L}_{int} = \int \mathcal{L} dt$ is defined as the integral of the luminosity with respect to time. The total number of events of one specific process is given by

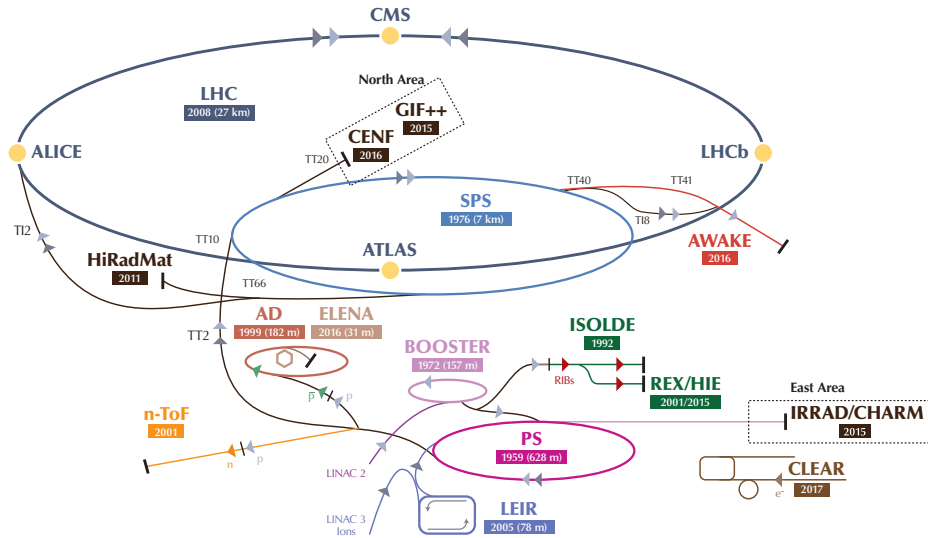


Figure 3.1.: The CERN accelerator complex. The proton bunches pass the pre-accelerators LINAC2, BOOSTER, PS, and SPS and then reach their final energy in the LHC with the four detectors ALICE, ATLAS, CMS, and LHCb. [56]

$$N = \mathcal{L}_{int} \sigma_p,$$

where σ_p is the corresponding cross section.

3.2. The ATLAS Detector

The ATLAS detector is a general-purpose detector at the LHC and has a cylindrical shape with a diameter of 25 m, a length of 46 m, and a weight of 7000 tons. It consists of three sub-systems: the tracking detector, the calorimeter system, and the muon spectrometer.

The detector is symmetric under rotations around the beam axis but divided into the barrel, end-cap, and forward region, due to different radiation characteristics in dependence on the polar angle. The detector components are arranged as co-linear cylinders around the beam axis in the barrel region and radially in the end-cap regions.

3.2.1. Coordinate System

The ATLAS detector employs a right-handed coordinate system with the origin at the collision point [57] as depicted in Figure 3.2. The x-axis points to the centre of the

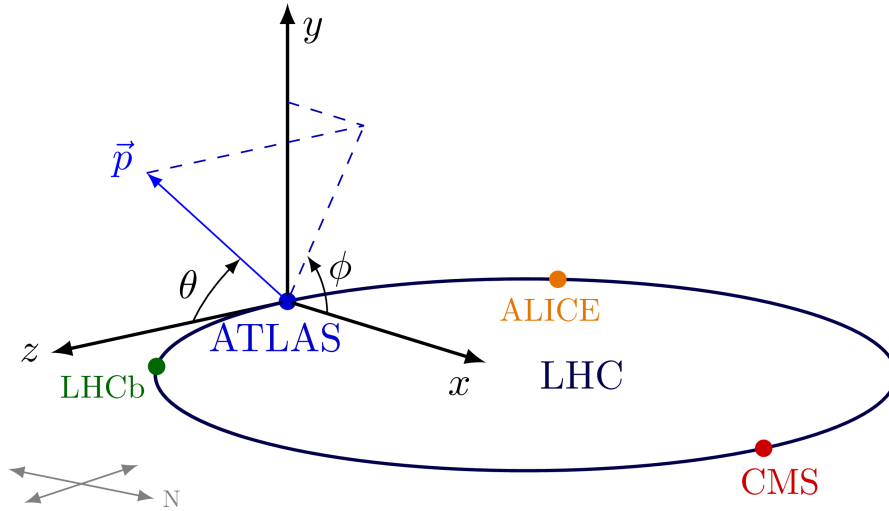


Figure 3.2.: The ATLAS Coordinate system.

LHC ring, the y-axis points upwards, and the z-axis is tangential to the accelerator ring. In order to describe the trajectories of particles, spherical coordinates are more suitable: the polar angle θ is measured around the x-axis, the azimuth angle ϕ around the beam-pipe, and the radius r denotes the distance from the coordinate origin.

Another quantity, the rapidity $y = \ln \left(\frac{E+p_L}{E-p_L} \right)$, is introduced, with the Energy E and the momentum parallel to the beam axis p_L . The rapidity y is preferred over the polar angle θ due to its invariance under Lorentz transformations. For highly relativistic particles (i.e. $m \ll p$), the rapidity converges to the pseudorapidity

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

Since not the protons, but their partons¹ interact, the momentum in the z-direction of the partons before the scattering process is unknown. Therefore, the transverse momentum

$$\mathbf{p}_T = \begin{pmatrix} p_x \\ p_y \end{pmatrix}$$

is a crucial observable since the transverse momentum before the scattering process is known to be zero. The absolute value of the transverse momentum is referred to as p_T . Further, the angular distance ΔR between two objects is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}.$$

3.2.2. The Inner Detector

The inner detector (ID) records space points of trespassing charged particles. Based on these space points, the trajectories (i.e. tracks), the event's primary vertex (i.e. the interaction point of the protons), and the decay position of unstable particles (i.e. the secondary vertices) are reconstructed. Since the track density decreases with increasing distance to the primary vertex, the ID is divided into three subcomponents, mainly differing by their resolution: the pixel detector, the semiconductor tracker (SCT),

¹The partons are the constituents of the proton, i.e. the valence quarks (uud), the gluons, as well as the quarks created by the gluon field.

3. Experimental Setup

and the transition radiation tracker (TRT). Each of the three components is installed differently in the barrel and the end-cap region of the detector. In the barrel region, the components are organised in concentric cylinders around the beam axis, while the components are set in disks perpendicular to the beam axis in the end-caps. The first and second components cover an η -range of up to 2.5, while the coverage of the outermost component is limited to $\eta < 2$. A cut-away view of the ATLAS ID system is shown in Figure 3.3.

The innermost of the three components is the pixel detector with an insertable b -layer (IBL). Each pixel has a size of $50 \times 400 \mu\text{m}^2$ in R - $\phi \times z$. The pixel sensors are arranged in three concentric layers. By the staggered arrangement of the pixels, an accuracy of up to $10 \mu\text{m}$ in R - ϕ and $115 \mu\text{m}$ in z is reached. The fine granularity is mainly motivated by enabling the identification of the secondary vertices of b -meson decays. The second component is the SCT, which operates similarly to the pixel detector but has lower granularity in the z -direction. The microstrip sensors of the SCT are arranged in four concentric double layers in the barrel region and nine layers in each forward region. The accuracy reached by the SCT is $17 \mu\text{m}$ in R - ϕ and $580 \mu\text{m}$ in z . In the pixel and SCT

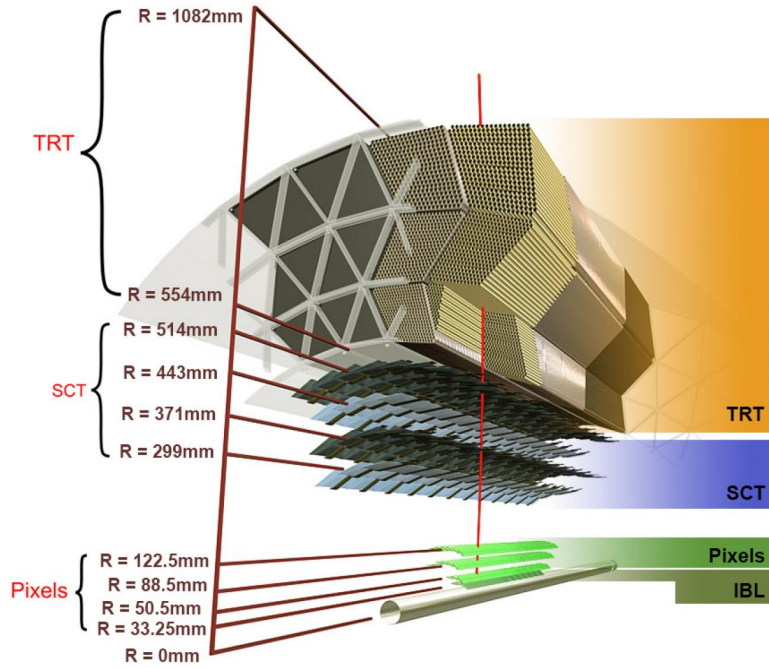


Figure 3.3.: Cut-away view of the ATLAS inner detector. [15]

detector, particles are measured using a semiconductor. The trespassing particles' excite the electrons of the scintillator from the valence to the conduction band. Then, the electrons and holes drift to the electrodes due to an external electric field and generate an electric signal.

The outermost component is a TRT made of gas-filled straw tubes. Trespassing particles ionise the gas, enabling free electrons to travel to the drift tube's centre and provide an electric signal. The straw tubes have an accuracy in the R - ϕ direction of $130\text{ }\mu\text{m}$. In addition to tracking, the detection of transition radiation allows differentiation between pions and electrons.

The ID is surrounded by a barrel-shaped superconducting magnet providing a solenoidal magnetic field parallel to the beam axis with a strength of 2 T. The magnetic field forces charged particles onto curved trajectories, from which the particles' momenta can be inferred.

3.2.3. The Calorimeter System

The calorimeter system is used to measure the total energy of electrons, photons, and hadrons and covers an $|\eta|$ -range of up to 4.9. The calorimeter system is subdivided into two parts: the inner part is the electromagnetic (EM) calorimeter (ECAL) which is used to measure the total energy of electrons and photons, whereas the hadronic calorimeter (HCAL) measures the energy of hadrons. The ATLAS calorimeters are sampling calorimeters with alternating layers of shower-inducing absorbers and active readout devices, as shown in Figure 3.4. The absorber materials' density increases from the barrel to the forward region as the total momentum of particles increases with η . The EM calorimeter has high granularity (i.e. small calorimeter cells) to measure the signal induced by electrons and photons with high precision.

Highly relativistic electrons and hadrons deposit energy predominantly via bremsstrahlung and photons via e^+/e^- pair production. If the energy of the bremsstrahlung photons is sufficient, they create an e^+/e^- pair. This process continues as a cascade until the energy of the photons is not sufficient for creating an e^+/e^- pair. These photons are then measured by inducing an electric signal in the active medium of the calorimeter. The radiation length X_0 is the mean distance travelled by an electron as its energy is reduced by $1/e$; λ is the mean free path between nuclear collisions. The ECAL (HCAL) has a length of more than $22X_0$ (9.7λ) in the barrel region and $24X_0$ (10λ) in the end-cap region. Hence, the calorimeter system provides a good containment of the particle showers and reduces the number of punch-through-jets².

²Punch-through jets are particles other than muons that reach the muon spectrometer.

3. Experimental Setup

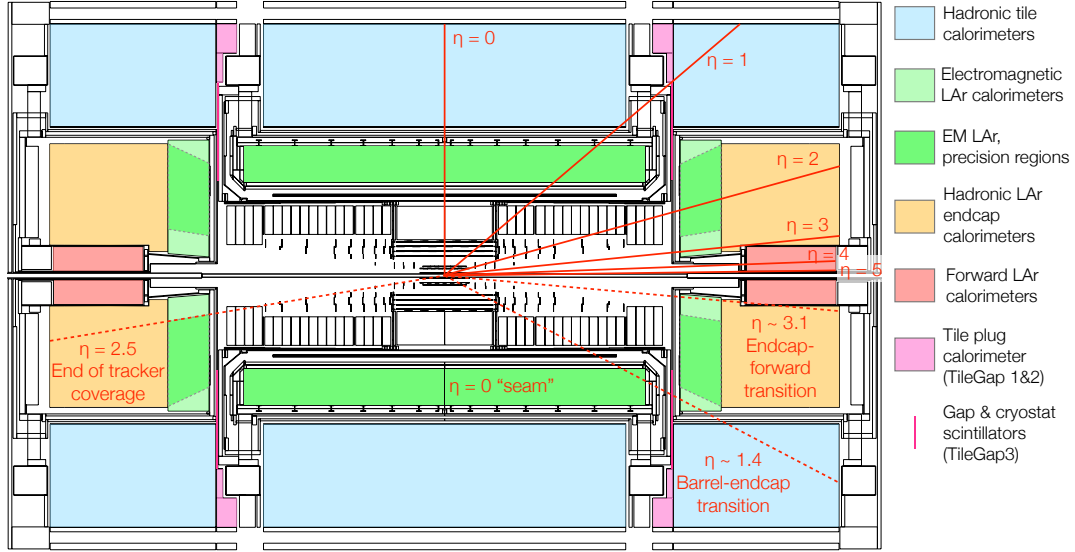


Figure 3.4.: A schematic view of the ATLAS calorimeter system. The EM barrel and end-cap calorimeters are shown in green. The EM end-cap has a precision region marked in darker green and an extended region in light green, and the transition from one to the other around $\eta \sim 2.5$ involves a significant change in the material layers. The hadronic Tile calorimeter is shown in light blue while the hadronic end-cap calorimeters based on liquid argon are illustrated in light orange. The forward calorimeters are shown in dark orange. [58]

The ECAL's active medium is made of liquid argon, and lead is used as the absorber material. The HCAL is built as a tile plug calorimeter and is used to measure the energy of hadrons. In the barrel region, the HCAL uses steel as an absorber and scintillator tiles as the active medium, while the hadronic end-cap calorimeter (HEC) uses liquid argon as its active medium, with copper as the absorber.

In the transition region ($\eta \in [1.32, 1.57]$) between the barrel and the end-cap, supply wiring of the tracking and the calorimeter system is installed, leading to degraded in energy resolution. Therefore, additional calorimeters (TileGap 1&2, TileGap 3) are placed in the transition region to counteract the degraded performance.

3.2.4. The Muon System

Muons have a significantly higher mass than electrons. Thus, they are minimum ionising particles (MIP), which means that they do not deposit a significant amount of their energy in the calorimeters and reach the muon spectrometer with almost their total initial momentum. The muon spectrometer is used to reconstruct muons; the underlying concept is similar to that of the ID: a strong magnetic field with 0.5 T (1 T) in the barrel (end-cap) region bends the paths of the trespassing muons. The momentum of the muon candidates is determined based on the deflection of the reconstructed tracks due to the toroidal magnetic field provided by the superconducting air-core magnets that surround the muon spectrometer.

In the range of η up to 2.7, muons are reconstructed using monitored drift tubes (MDT). The MDTs are gas-filled and have a wire spanning in the centre. When muons enter these tubes, they ionise the gas, and the free electrons drift to the wire and form an electric signal due to the voltage between the wire and the tube.

Due to their faster response, cathode strip chambers (CSC) are employed in the muon system's innermost module ($\eta \in [2, 2.7]$). Besides the muon spectrometer, resistive plate

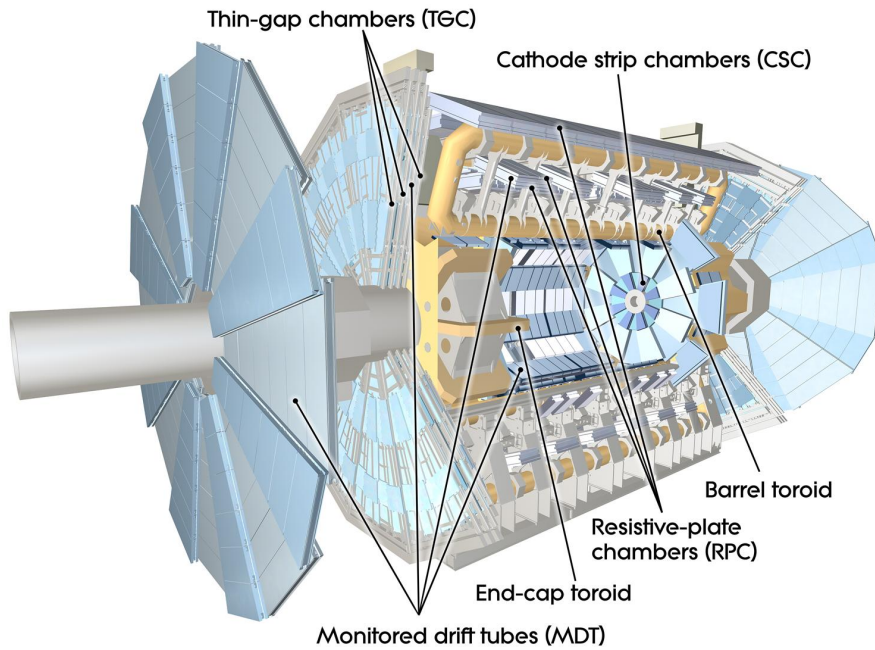


Figure 3.5.: Cut-away view of the ATLAS muon system. [15]

chambers (RPCs) and thin gap chambers (TGCs) are installed in the barrel and end-cap regions for the trigger system (see Section 3.2.5). A cut-away view of the ATLAS muon system is provided in Figure 3.5. [59]

3.2.5. The Trigger and Data Acquisition System

At the ATLAS detector, bunches cross with a frequency of 40 GHz. Since the research program at the LHC aims to study nature at the smallest distances and the highest energies, soft QCD events are not of interest. The trigger system filters potentially relevant events to reduce the number of events to a processable size.

It consists of two stages: The Level 1 (L1) trigger is a hardware trigger based on information from the calorimeter system's reduced granularity regions and the muon system's RPC and TGC. Events are filtered based on information on the transverse jet energy, the missing transverse energy, and the η/ϕ region. The L1 trigger reduces the event rate to approximately 100 kHz with a latency of 2.5 μ s.

The second trigger stage is a software-based high-level trigger (HLT) that filters events based on reconstructed physics objects similar to the objects processed in the offline analysis. Based on these properties, the HLT filters reduce the event rate to 1.2 kHz generating 1.2 GB/s with a latency of 40 ms. These events are then saved to the permanent memory for offline analysis. [60, 61]

3.3. Reconstruction and Identification of Physic Objects

Powerful event reconstruction algorithms transform raw detector information into analysis objects to infer physical particles. These analysis objects are trajectories and topological clusters: hits in the inner detector and the muon spectrometer form trajectories, and topological clusters are reconstructed from energy deposits in the calorimeter system. The following section provides an overview of the jet clustering and particle reconstruction algorithms at ATLAS. [62]

3.3.1. Electrons

Figure 3.6 presents the standard electron identification and reconstruction procedure [31, 63, 64]; it is based on the reconstruction of clusters in the EM calorimeter, tracks in the ID and the matching of these two objects.

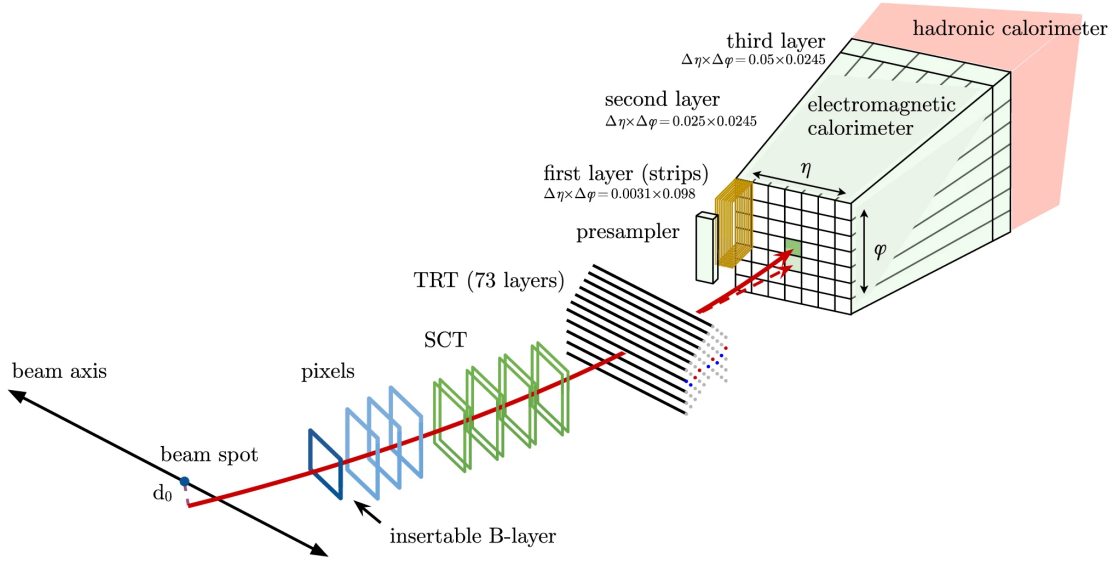


Figure 3.6.: An illustration of an electron trespassing the detector. The red trajectory shows the path of an electron, traversing the tracking system (pixel detectors, silicon-strip detectors, TRT) and then entering the electromagnetic calorimeter. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron. [63]

Track Reconstruction The reconstruction algorithm for particle trajectories processes the ID measurement. It consists of three steps: creating space points from raw ID hits, fitting the space points to tracks, and refitting the tracks. These steps are outlined in the following paragraphs and more detail is given in Reference [31].

At first, space points are reconstructed from clusters of hits in the ID [65]. Then, three space points in the silicon detector serve as a seed for the track reconstruction algorithm. Next, the baseline algorithm tries to extend the seed to a track with a pion hypothesis to model the energy deposition of the trespassing particle in the ID. Finally, a subsequent track-fitting attempt is applied to seeds with $p_T > 1$ GeV that could not be extended to a full track if the EM cluster meets specific shower shape requirements. This attempt accounts for energy losses of up to 30 % at every intersection with the detector material due to bremsstrahlung.

Then, the ATLAS Global χ^2 Track Fitter [66] is applied to the preliminary tracks with $p_T > 400$ MeV and resolves ambiguity between hits associated with multiple tracks. If the Track Fitter fails under the pion hypothesis and the separation of the track and the associated EM cluster fulfils specific criteria. The refitting is repeated with an additional

Gaussian term accounting for bremsstrahlung.

After the extension of the seeds and the preliminary refitting, a subsequent refitting algorithm, based on a Gaussian sum filter (GSF) [67], is applied to the raw detector measurement of tracks with at least four silicon hits loosely matched to an EM cluster. This method is based on a superposition of weighted Gaussian functions and accounts for non-linear energy deposition caused by bremsstrahlung. The Gaussian functions describe two effects - six functions characterise the energy loss in the material and up to twelve functions track parameters.

Topological clusters The $\eta \times \phi$ space of the electromagnetic (EM) calorimeter is divided into 200×256 elements, corresponding to the granularity of the second layer of the EM calorimeter ($\Delta\eta \times \Delta\phi = 0.025 \times 0.025$). For each element, the energy of the first three components of the EM calorimeter is summed. These objects define a tower. A cluster candidate is seeded if the energy in a window of 3×5 elements exceeds the energy of 2.5 GeV. The window slides over the $\eta \times \phi$ grid with a step size of 0.025 in both directions [68]. If the separation of two seed cluster candidates is less than 5×9 elements, one of the two seeds is discarded. If the energy difference between the two clusters is above ten percent, the seed with the lower energy is discarded; in the other case, the cluster in the central tower with the greater energy is preserved.

Track Cluster Matching After reconstructing the GSF tracks and calorimeter-seed clusters, matching the two objects completes the electron candidate reconstruction. The track-cluster matching is similar to the matching before the GSF refitting but with stricter requirements in the azimuthal separation. The azimuthal separation of the cluster position and the track has to fulfill the requirement $-0.10 < \Delta\phi < 0.05$; and the cluster position and the track with rescaled momentum to match the cluster energy has to fulfill the requirement $-0.10 < \Delta\phi_{res} < 0.05$. The ambiguity of clusters matched to more than one track is resolved based on the distance between the extrapolated tracks and the clusters in the second component of the calorimeter, and the number of hits in the silicon detectors. An electron candidate is defined if the associated track has at least four hits in the silicon layers and is not associated with a vertex from a photon conversion.

E/γ supercluster For electron reconstruction, variable-sized superclusters are formed from the topological clusters. The supercluster reconstruction procedure consists of two stages. First, a seed cluster is selected from a list of EM topo-clusters. Then, EM topological clusters near the seed are identified as satellite clusters, which may arise

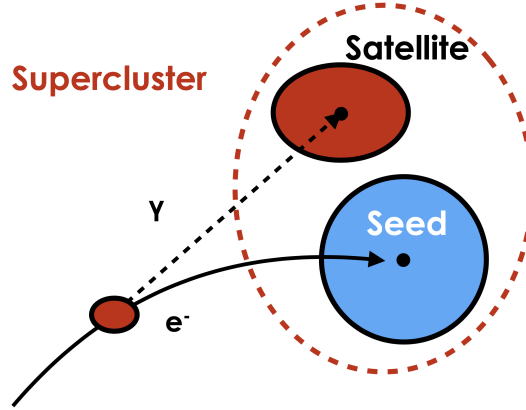


Figure 3.7.: Supercluster with a seed electron cluster and a satellite photon cluster. [64]

from bremsstrahlung radiation or the splitting of topological clusters. If the satellite clusters satisfy the necessary selection criteria, they are added to the seed cluster to form the final supercluster as shown in Figure 3.7.

Identification The standard electron identification at ATLAS is based on a likelihood method that combines information on electromagnetic shower shapes, track quality, and track-cluster matching in order to reduce backgrounds. The Likelihood (LH) for signal and background is based on the product sum of the probability density functions (PDF) P_j with $j \in \{S, B\}$ of n observables. The PDFs for the electron LH are derived from simulation samples. First, separate PDFs binned in E_T and η are determined for each observable to account for the kinematic and spatial dependence of the input quantities. Then, the PDFs are generated from finely binned histograms of the individual identification observables. The Likelihood is defined as:

$$L_{S(B)}(\mathbf{x}) = \prod_{i=1}^n P_{S(B),i}(x_i), \quad (3.2)$$

with the vector $\mathbf{x}$ of identification observables of an electron candidate. Based on each electron candidate's signal and background LH, the discriminant

$$d_L = \frac{L_S}{L_S + L_B} \quad (3.3)$$

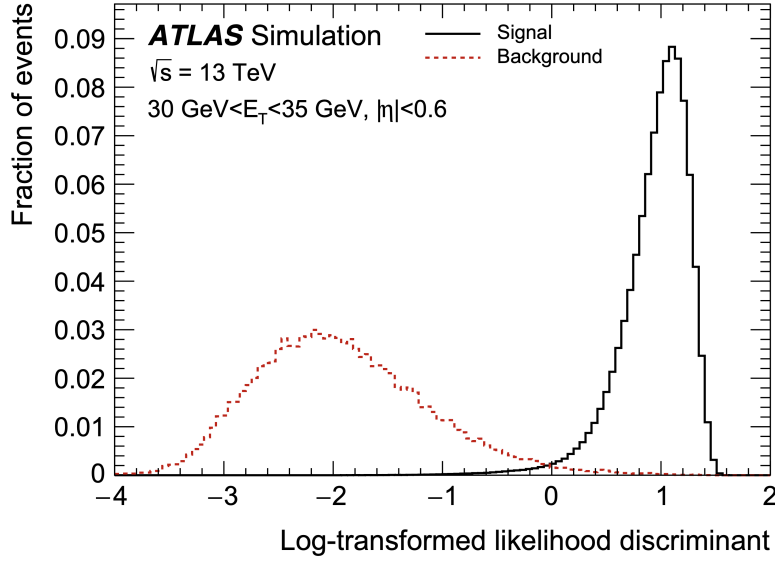


Figure 3.8.: The transformed discriminant d'_L for reconstructed electron candidates. The black histogram is for prompt electrons, and the red line is for the sum of backgrounds. [63]

is determined. The discriminant d_L peaks at one (zero) for signal (background). Selecting thresholds to distinguish between signal and background is inconvenient due to the sharp peaks. Therefore, the inverse sigmoid function is applied to the discriminant to increase the resolution in the region with overlapping of the signal and background distributions.

$$d'_L = -\frac{\ln(d_L^{-1} - 1)}{15} \quad (3.4)$$

The separation power of the discriminant can be seen in the histogram Figure 3.8. The *Loose*, *Medium*, and *Tight* operating points are defined with efficiencies for identifying a prompt electron with $E_T = 40$ GeV of 93 %, 88 %, and 80 %, respectively. The operating points correspond to increasing thresholds for the LH discriminant. Electron candidates with values of d'_L larger (lower) than the operating point are classified as signal (background). [63]

Electron isolation Electron candidates must fulfil isolation criteria to suppress the contamination by electrons from heavy-flavour decays and misidentified hadrons. Signal electrons are characterised by little activity in their vicinity. The isolation

variables are constructed both on calorimeter and ID information. For the calorimeter-isolation variable $E_{T,\text{cone}}^{\text{isol}}$, the energy depositions within a cone of radius ΔR around the electron candidate are summed. Then, the energy within a rectangle of size $\Delta\eta \times \Delta\phi = 0.125 \times 0.125$ in the centre of the cone is subtracted to account for the energy of the electron candidate. Respectively, for the track-isolation variable p_T^{isol} , the momenta of tracks with $p_T > 1$ GeV within a cone ΔR around the electron candidate are summed, and the candidate's momentum subtracted. [63]

3.3.2. Muons

The measured space points in the muon detector are reconstructed using algorithms similar to those for ID tracks. Depending on the matching of the measurement in the three detector components (i.e. the ID, the calorimeter and the muon spectrometer), the muon candidates are assigned to four classes:

1. *Combined* muons are built from a track in the ID matched to a track in the muon spectrometer.
2. *Segment-tagged* muons are built from an ID track that could not be matched to a full track in the muon spectrometer but to a local track segment.
3. *Calorimeter-tagged* muons are ID tracks matched to an energy cluster in the calorimeter system.
4. *Extrapolated muons* are built from muon spectrometer tracks and need to be compatible with the primary scattering vertex.

Muons are identified based on selection requirements on the number of hits in the tracking system (i.e. the ID and muon spectrometer), the track fit properties, and the compatibility of the measurements in the three detector components. For each of the four muon types, three sets of requirements are defined as the *Loose*, *Medium* and *Tight* working points with increasing purity and decreasing efficiency. Muons with very high- p_T have almost straight trajectories, and the matching of the measurement in the three detector components becomes the limiting factor. To account for these differing requirements, the High- p_T working point is defined. [59, 69]

3.3.3. Jet Clustering

As a result of colour confinement, quarks cannot exist isolated but confined in colourless bound states. Combining partons with colour and anti-colour or all three (anti-)colours forms a colourless state called hadron [70]. Quarks produced in particle collision induce

the so-called particle shower. These hadronic particle showers are reconstructed via jets.

Topological Clustering The topological clustering algorithm groups energy deposition in neighbouring calorimeter cells into groups, the so-called calorimeter clusters. At first, calorimeter cells with energies whose signal-to-noise ratio (S/N) exceeds a value of four are identified as seeds. Secondly, cells with $S/N > 2$ are added to the seeds. Thirdly, all neighbouring calorimeter cells are added, and the topological cluster is complete. Then, the energy of the topological cluster is the sum of all its calorimeter cells. Finally, the location of the topological clusters is defined as the location of all associated calorimeter cells weighted by the respective cell energies. [71, 72]

Particle Flow The particle flow (PFLOW) algorithm [72] improves the reconstruction of physic objects by combining the ID and calorimeter system's advantages; it prevents double counting of energy clusters and isolates overlapping particles.

Particles may deposit energy in multiple topological clusters. Therefore, the PFLOW algorithm combines topological clusters potentially stemming from the same particle into single-clusters based on a probability metric. Particle Flow objects are assigned to two classes: charged particle flow objects are ID tracks matched to the primary vertex, while neutral particle flow objects are modified topo-clusters in the calorimeter system. With descending transverse momentum of the tracks, tracks are matched to the cells of the calorimeter. The expected energy based on the particle's trajectory is subtracted from the energy deposited in the calorimeter cell.

The PFLOW algorithm proceeds in several steps: At first, muons are reconstructed, which generate a signal in the ID and muon detectors; corresponding signatures are subtracted. Then, the reconstruction of electrons follows. By extrapolating the track to the ECAL, the corresponding energy clusters are subtracted from the event's signature. Subsequently, charged hadrons are identified. The charged hadrons create a trajectory in the ID and clusters in the calorimeters. If momentum, energy, η , and ϕ are compatible, the PFLOW algorithm assigns the measurement of the two subdetectors to one particle. After reconstructing all particles with signal in the ID, the neutral hadrons and photons are reconstructed. Photons are reconstructed from clusters in the ECAL; neutral hadrons form clusters in the HCAL.

Jet Clustering Algorithms Jets clustering algorithms are divided into the classes cone based and iterative. Cone-based algorithms maximise the in-circle energy, while

iterative algorithms repeatedly form pair-wise combinations of four-momenta and are used exclusively at the LHC.

The Iterative algorithms described below are infrared and collinear (IRC) safe. A flowchart visualising the iterative jet clustering algorithm is provided in Figure 3.9. Infrared safe algorithms are not changing with soft emissions of the final-state particles [73]. Collinear safety refers to the stability of the jet definition when splitting a final-state particle into two collinear components. Iterative algorithms recombine the closest pairs of particles based on a specific metric. The distance metric used at ATLAS is given by

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}$$

$$d_{iB} = k_{ti}^{2p}$$

with transverse momentum k_t of particle i and j and the angular separation $\Delta_{ij} = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$. The radius parameter R controls the cone size, and p governs the relative influence of geometrical and energy dependence on the metric.

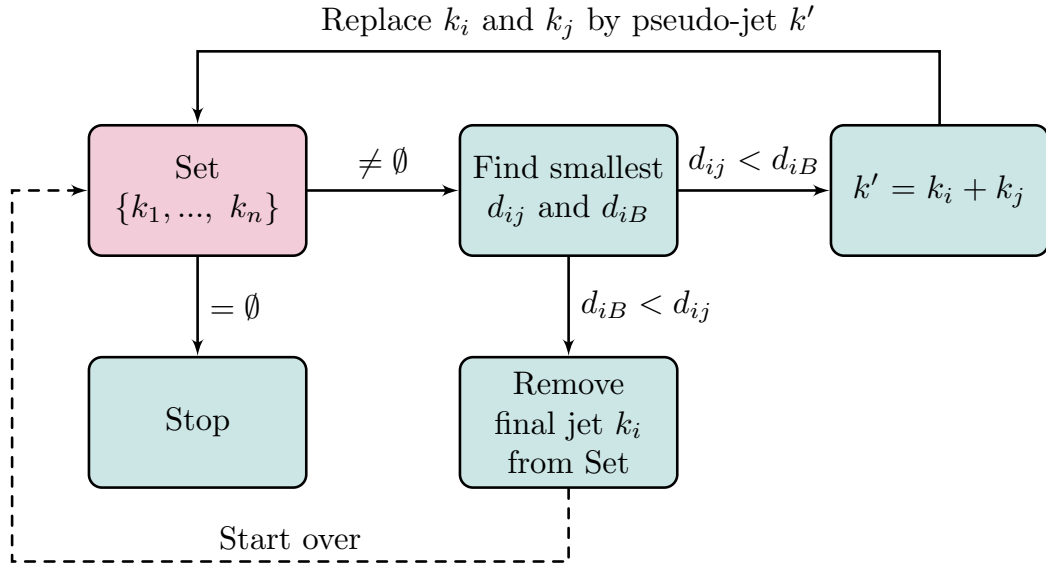


Figure 3.9.: Representation of iterative jet clustering algorithms as a flowchart. The starting point is the red node corresponding to the set of four-momenta of clusters created by a clustering algorithm.

The components needed for the jet clustering algorithm are complete: the four-momenta of the individual particles and a distance measure. Additionally, well-defined instructions must be provided for combining the individual four-momenta into jets.

At first, the smallest d_{ij} and d_{iB} of the full set of four-momenta (k_t) are determined. If the smallest distance is any $d_{ij} < d_{iB}$, the clusters p_i and p_j are replaced by the four-vector sum of the two elements, the so-called *pseudo jet*. However, if $d_{iB} < d_{ij}$ and p_i is called a *final jet*. This procedure is repeated until no constituents or pseudo-jets remain in the initial set.

Depending on the exponent p , the algorithm is called k_t ($p = 1$), *anti- k_t* ($p = -1$), or *Cambridge/Aachen* ($p = 0$). The inter-particle distance of Cambridge/Aachen is purely geometrical; for k_t (anti- k_t), it depends on the inverse transverse momentum of the harder (softer) particle. The anti- k_t algorithm [74] clusters hard energy depositions first, followed by soft clusters. Hard clusters, therefore, influence energetic cores primarily, and after replacing the hard clusters with pseudo jets, almost perfectly conical jets are produced by adding the soft clusters.

The default jet clustering algorithm at ATLAS is the anti- k_t algorithm with a radius parameter of 0.4 for small-Radius jets and 1 for large-Radius jets.

Vertices Collision vertices are reconstructed from at least two ID tracks, each with a transverse momentum $p_T > 500$ MeV [75]. Among all vertices, the one with the highest p_T^2 sum of all associated tracks is chosen to be the primary vertex (PV) of the event. The properties of charged particle tracks in the ID are calculated with respect to this PV.

b-Tagging b -quarks are of special interest for the research at ATLAS as they often originate from events including Higgs bosons or t -quarks. Background contamination by light-flavored jets is reduced by employing the unique properties of jets induced by b -quarks. The decay of the b -meson is CKM suppressed; therefore, b -mesons have a relatively long mean lifetime of $\tau \sim 1.6$ ps [76]. This unique feature is used to identify events, including b -quarks (see Figure 3.10). Due to the high granularity of the pixel detector, the secondary of the b -meson decay can be resolved. Furthermore, b -quarks are the second-heaviest quarks in the SM and decay to the significantly lighter charm and up quarks; the mass is converted into kinetic energy leading to a large opening angle of the jet cone. Hence, tracks emerging from b -hadrons have a significant longitudinal and transverse impact parameter.

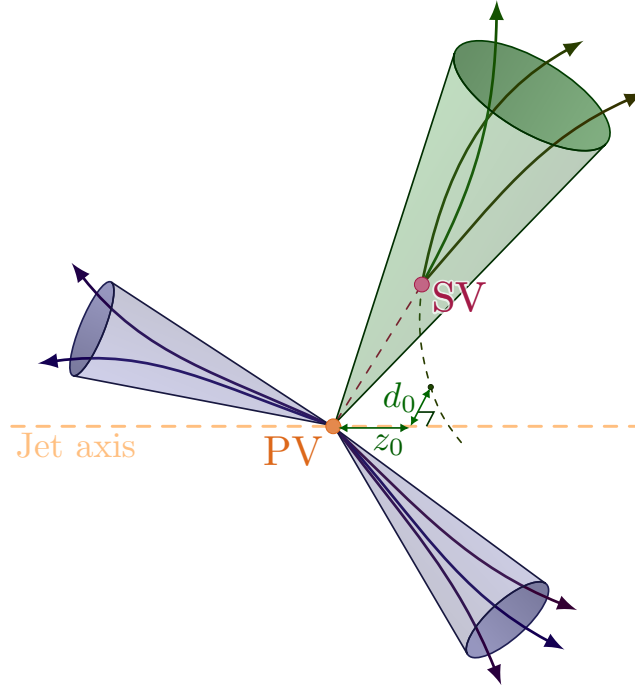


Figure 3.10.: Visualisation of the unique signature of bottom quarks in comparison to light quarks. In contrast to light quarks, the bottom quarks form a secondary vertex.

B -tagging identification is divided into two steps: Initially, low-level algorithms process observables, such as the impact parameter of trajectories in the ID associated with jets. Algorithms used at ATLAS are the Likelihood-based algorithms IP2D and IP3D and the RNNIP tagger based on a Recurrent Neural Network [77]. In the second stage, the output of the low-level tagger is processed by high-level taggers. High-level taggers at ATLAS are the Boosted Decision Tree (BDT) based tagger MV2 and the Deep Neural Network DL1 [78]. Figure 3.11 shows the background rejection in dependence of the b -jet efficiency for the MV2c10 algorithm, where c10 indicates the c -jet fraction in the training sample of 10 %.

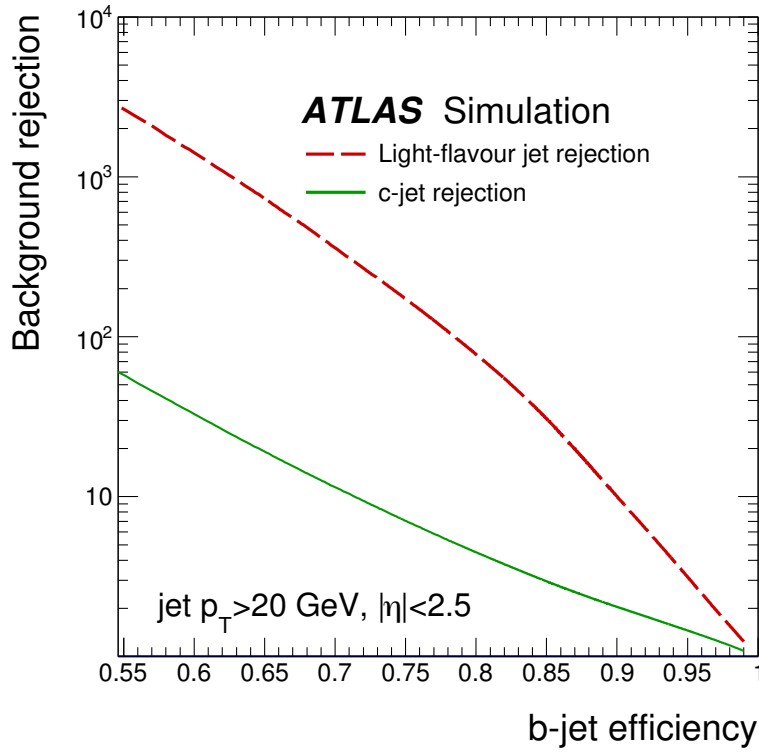


Figure 3.11.: The light-flavour jet (dashed line) and c -jet rejection factors (solid line) as a function of the b -jet tagging efficiency of the MV2c10 b -tagging algorithm. [79]

3.3.4. Jet Energy Correction

In the following, an overview of the calibration techniques of clusters in the calorimeter system and different jet collections is provided (see Figure 3.13).

The jet correction at ATLAS is divided into two main steps: first, corrections are applied to the clusters, the so-called constituent level correction. Then, after creating jets of the individual clusters with a jet clustering algorithm, the jet energy scale (JES) corrections are applied to the clustered jets.

Initially, the clusters in the calorimeters are calibrated to respond equally to electromagnetic showers induced by photons and electrons. In addition to the EM calibration, the so-called local cluster (LC) weighting can be applied to the clusters. The LC weighting intends to calibrate the calorimeter signal without taking any concrete assumption about the particle type. The response of HAD objects in the calorimeter is significantly lower than the response to electromagnetic objects since the energy deposited via the strong force provides no electrical signal in the detector system. Hence, calorimeter clusters induced by hadronic objects require tuning with significantly larger magnitudes than electromagnetic objects [71]. The LC weighting scheme incorporates this feature by determining whether the cluster was induced by an EM particle ($\mathcal{P}_{\text{clus}}^{\text{EM}}$) or a HAD particle ($1 - \mathcal{P}_{\text{clus}}^{\text{EM}}$). The weight applied to a single cell is composed of the scale factors for corrections at the EM ($w_{\text{cell}}^{\text{em-cal}}$) and HAD scale ($w_{\text{cell}}^{\text{had-cal}}$) multiplied by the corresponding probabilities.

$$w_{\text{cell}}^{\text{cal}} = \mathcal{P}_{\text{clus}}^{\text{EM}} \cdot w_{\text{cell}}^{\text{em-cal}} + (1 - \mathcal{P}_{\text{clus}}^{\text{EM}}) \cdot w_{\text{cell}}^{\text{had-cal}}$$

Besides balancing the response to EM and HAD objects, the LC weighting scheme applies out-of-cluster and dead-material corrections.

Then, a clustering algorithm like anti- k_t combines the calibrated clusters into jets. Finally, the so-called jet energy scale (JES) correction is applied to the jets [80, 81]. First, the jet direction is corrected from the detector centre to the hard-scatter vertex. Second, the expected pile-up contribution is subtracted from the jets. Third, Monte Carlo (MC) derived η and energy-dependent corrections are applied to the reconstructed four-momentum. Third, Monte Carlo (MC) derived η and energy-dependent corrections are applied to the reconstructed four-momentum. Therefore, the response distribution

$$\mathcal{R} = \frac{E^{\text{reco}}}{E^{\text{truth}}}, \quad (3.5)$$

with the energy of the reconstructed jet E^{reco} and the truth level jet E^{truth} , is determined. Reconstructed jets are geometrically matched to the truth-level jets via a ΔR criterion.

The average response is defined as the mean of a Gaussian fit applied to $\mathcal{R}$. The jet energy calibration factor is determined as the inverse of the average response parametrized in η and E^{truth} . Figure 3.12 depicts the η dependence of the average response: the average response is mitigated the barrel-endcap ($|\eta| \sim 1.4$) and the endcap-forward ($|\eta| \sim 3.1$) transition region due to different calorimeter technologies and change in calorimeter granularity. [82, 83].

A conclusion of the jet energy correction at ATLAS is provided in Figure 3.13.

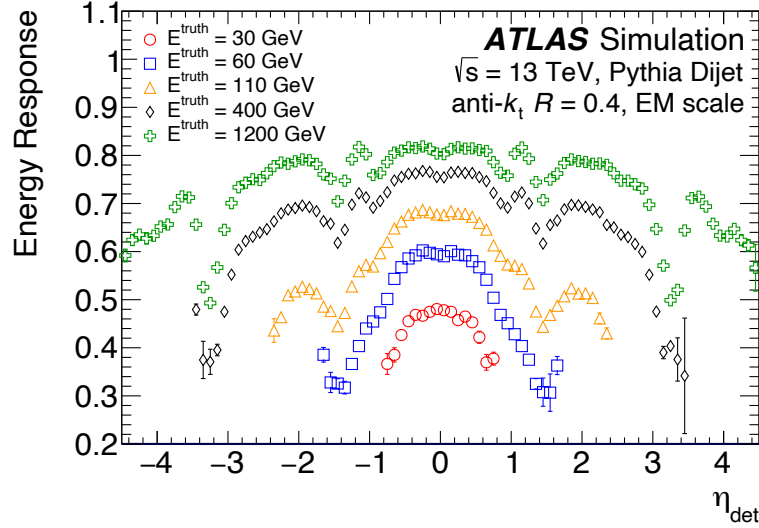


Figure 3.12.: The average energy response as a function of η for different energies. The energy response is shown after origin and pile-up corrections are applied. [83]

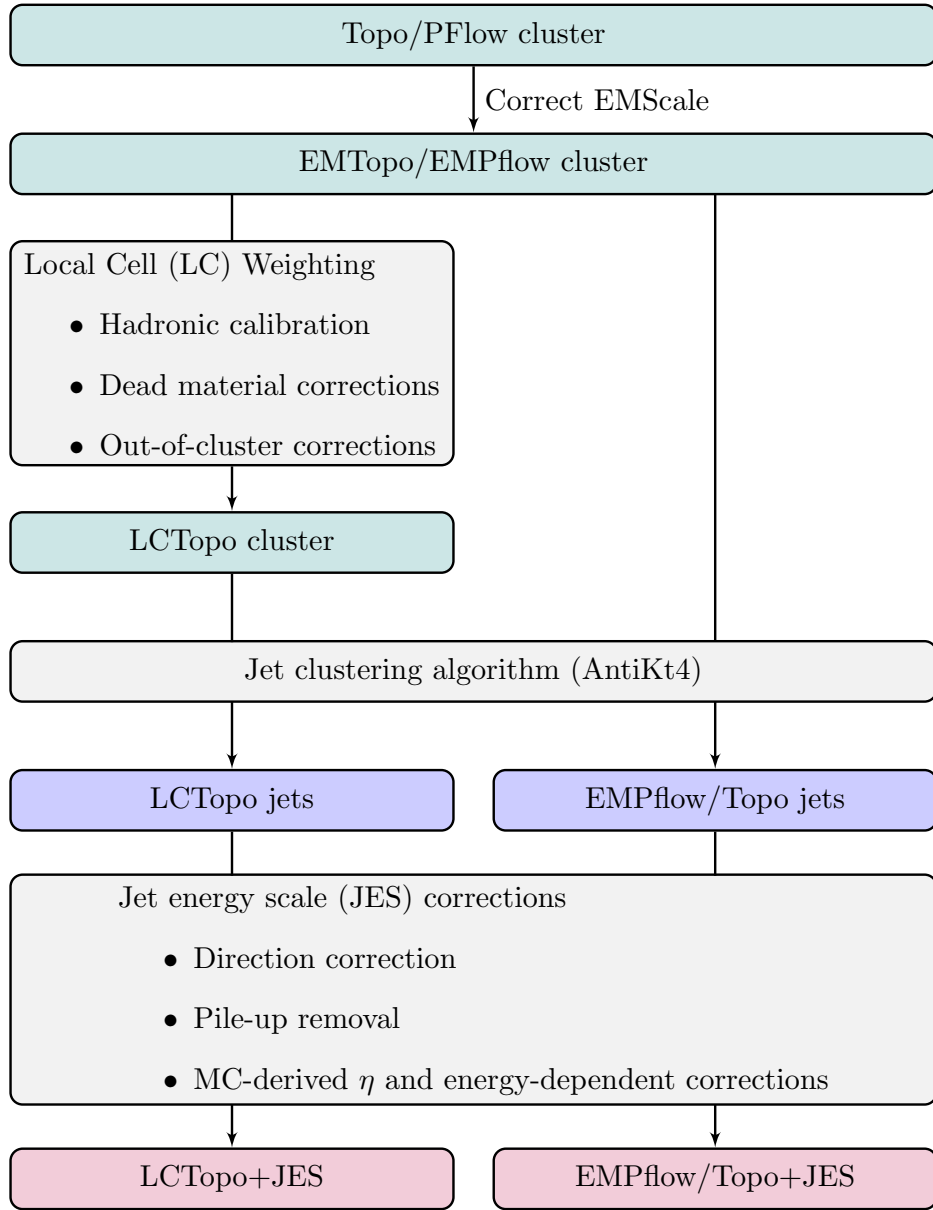


Figure 3.13.: Overview of the cluster and jet collections at ATLAS. Clusters are shown in green, constituent-level jets in blue, and fully calibrated jets in red. The algorithms transforming the different objects are outlined in the grey nodes.

3.3.5. Missing Transverse Energy

Neutrinos can not directly be reconstructed with the ATLAS detector. In order to compensate for this deficiency, the vectorial sum of all particles' transverse momenta is calculated. Due to momentum conservation, the transverse momenta of all objects in the process sums up to zero. Missing transverse energy (MET) hints at particles that are not detected by the ATLAS detector like neutrinos, which only interact weakly. The MET is defined as

$$E_T^{\text{miss}} = -\sum p_T^{\text{hard}} - \sum p_T^{\text{soft}}, \quad (3.6)$$

with the transverse momenta of the fully reconstructed and calibrated particles (i.e. the hard objects) p_T^{hard} and the transverse momenta of charged particle tracks that are associated with the hard-scattering vertex (i.e. the soft objects) p_T^{soft} . [84–86]

4. Introduction to Deep Learning

Deep learning has seen rapid growth in the last decade, and its potential applications have led to hardware advancements, new workhorse algorithms, and user-friendly software. The scope of applications includes natural language processing over autonomous driving to particle physics analysis. Deep Learning is one branch of Machine Learning which can be defined as “the study of computer algorithms that allow computer programs to automatically improve through experience” [87].

Supervised and unsupervised learning are two main fields of deep learning. In supervised learning, the desired prediction of the deep neural network (DNN) is well-defined (ground truth). During the training, a predefined loss function evaluates the network’s performance by comparing the prediction to the ground truth. Then, in the optimisation step, the optimiser adapts the DNN’s parameters to minimise the loss as depicted in Figure 4.1. In contrast, unsupervised learning extracts previously unknown patterns and correlations from unlabeled data.

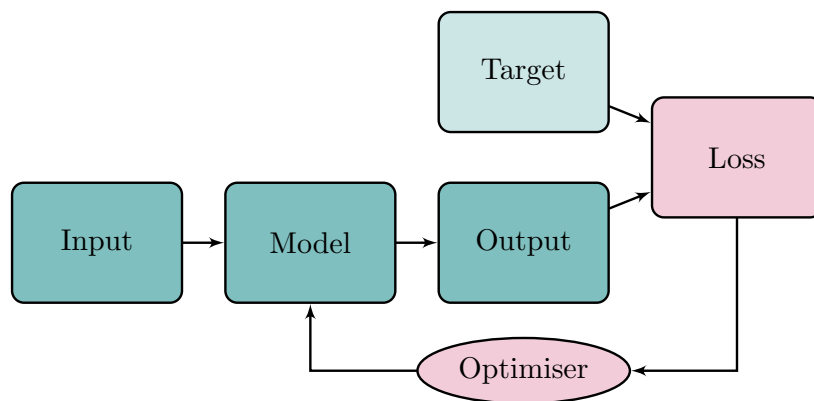


Figure 4.1.: Visualisation of the underlying concept of supervised learning. The input is fed into the model, which outputs a prediction. Target, loss, and optimiser govern the training procedure; the loss function compares the target and prediction, and the optimiser adjusts the model for an improved prediction with respect to the loss.

Furthermore, the applications of DNNs are divided into classification and regression tasks: in classification tasks, the output is a vector where every entry represents the probability of the input belonging to the corresponding class, e.g., the particle's type based on its signature in the ATLAS detector. In contrast, regression tasks predict continuous values such as momentum, temperature, or age.

4.1. The Artificial Neuron

The artificial neuron, or perceptron, is the basic building block of a DNN. Like a biological neuron, it maps an input over an activation to an output signal; it can be described as a mathematical function containing a set of weights w_i , a bias b , and a non-linear activation function f . An illustration of an artificial neuron is provided in Figure 4.2. The principle is the following: the inputs of the neuron x_i are weighted and summed (similar to the inner product on the euclidean space) plus the bias and afterwards passed on to the activation function. Within this mathematical conception, the bias represents the stimulus threshold, and every weight quantifies the strength of each connection.

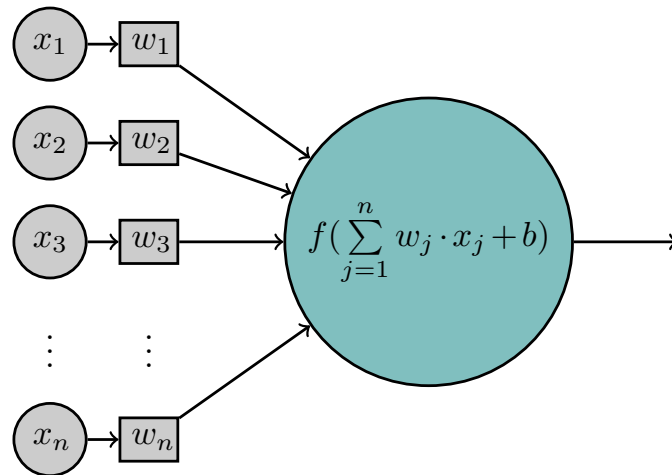


Figure 4.2.: Illustration of an artificial neuron: each component x_i of the input is weighted by a factor w_i . Then, the sum is determined and a bias b is added, before applying the activation function f .

Activation Functions The activation function is applied to the weighted sum of inputs of a neuron. To enable the representation of non-linear behaviour, activation functions are used. In addition to their non-linearity, activation functions should be differentiable with low computational effort. Two common activations are the sigmoid and rectified linear units (ReLU).

The ReLU function is defined as

$$f : \mathbb{R} \rightarrow (0, \infty), x \mapsto \max(0, x)$$

$$\frac{d}{dx}f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}.$$

A shortcoming of ReLU nodes is that they can *die*, then the neuron's output is zero for every sample, and the node will no longer contribute to the classification process. Another activation function is the sigmoid function which maps the input to the range $[0, 1]$. The differentiation is computationally efficient since its gradient is expressible by the sigmoid function itself.

$$\sigma : \mathbb{R} \rightarrow (0, 1), x \mapsto \frac{1}{1 + e^{-x}}$$
$$\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))$$

The sigmoid function suffers from the vanishing gradient problem, which can cause the network to learn slowly or not at all since the gradient converges to zero for $|x| \rightarrow \infty$. As a result, only very slight changes are made to the parameters. The ReLU and Sigmoid activation functions are shown in Figure 4.3.

4.2. Architectures

Perceptrons are the basic building blocks and their arrangement defines the DNN's architecture. The nodes in a feed-forward neural network are arranged in layers, and the nodes are connected precisely to the nodes in the next layer. The first (last) layer is called the input (output) layer and contains neurons according to the number of input

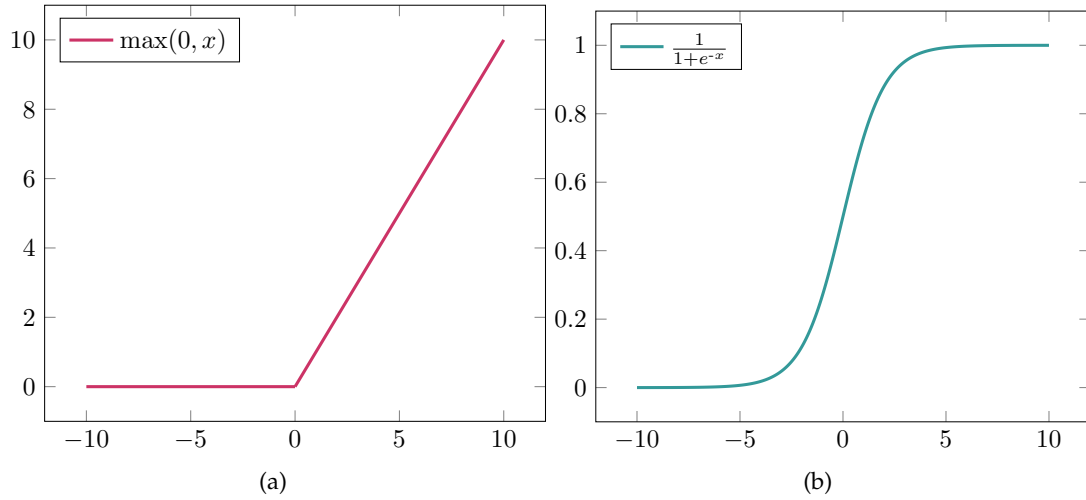


Figure 4.3.: The (a) ReLU and (b) Sigmoid activation function.

parameters (i.e. the number of classes). The layers in between are called hidden layers which determine the complexity of the network. The term deep neural network is not well-defined but refers to the layered structure. Usually, a neural network with at least two hidden layers is considered to be *deep*.

Each layer is a vector with entries corresponding to the number of nodes in the respective layer. The mapping from one layer to the subsequent layer can be thought of as a simple linear function. Neuronal connections can be arranged in i matrices with entries w_{ijk} (weights); their position in the DNN is well-defined by the layer i and the position j (k) in the first (subsequent) layer. Furthermore, to each weighted sum of inputs of a node, the bias b_{ij} is added (see Figure 4.4). In the following, all model parameters w_{ijk} and b_{ij} are summarised in the vector $\theta = \{w, b\}$. The network's modular structure offers high flexibility in the DNN's complexity by adjusting the number of hidden layers and neurons.

The structure of a DNN can be exemplified by a network for recognising single handwritten numbers between 0 and 9. The input layer's neuron count corresponds to the image's pixel count representing the handwritten number. The DNN maps the input parameters to the output layer via the hidden layers. The output layer consists of ten neurons - one for every digit from 0 and 9. The value range of each output node is the interval $[0, 1]$ and is associated with the probability to be the corresponding number.

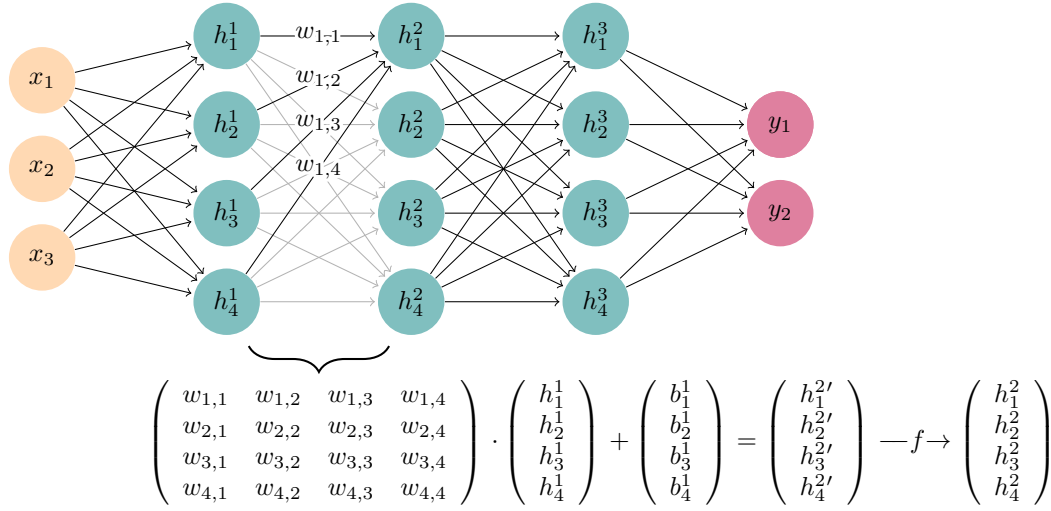


Figure 4.4.: Illustration of a DNN as an ensemble of connected nodes and the analogue matrix representation. The mapping of one layer to the subsequent layer can be thought of as matrix multiplication. The bias vector $\mathbf{b}$ represents the stimulus threshold of each node in a layer. The non-linear activation functions f provides additional complexity.

4.3. Training

A DNN is trained by comparing its prediction with the ground truth and then modifying its parameters to improve the prediction. The parameters are initialised randomly before training; therefore, the DNN's performance at this stage corresponds to random guessing.

Loss functions

The loss function is a metric that quantifies the goodness of the classifier by evaluating its prediction concerning the ground truth. The mean squared error (MSE) loss

$$L_{MSE}(\mathbf{y}, \hat{\mathbf{y}}(\boldsymbol{\theta})) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i(\boldsymbol{\theta}))^2, \quad (4.1)$$

is a widely used loss function for binary classification tasks with the ground truth y_i

and the prediction of the DNN $\hat{y}_i(\theta)$ in dependence on the network parameters θ . The loss value is divided by the number of samples n .

For multi-class tagging, the categorical cross-entropy is well suited: before applying cross-entropy loss, the output layer must be converted into a probability-like object; thus, the SoftMax function

$$\hat{y}_i \rightarrow \text{SoftMax}(\hat{y}_i) = \frac{\exp(\hat{y}_i)}{\sum_{j=1}^k \exp(\hat{y}_j)}$$

is applied to the output layer, with the value y_i the i^{th} output node. The SoftMax function ensures that the values in the output layer lie in the range $[0, 1]$ and sum up to one. Then, the cross-entropy loss is applied to the transformed output vector, which is a measure of the difference between two probability distributions for a given random variable or set of events and is well-suited as a loss function for multi-class DNNs:

$$L_{CE}(\mathbf{y}, \hat{\mathbf{y}}(\theta)) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k (y_{ij} \cdot \log \hat{y}_{ij}) .$$

In addition to the indices introduced in Equation 4.1, k is the number of classes. Using the CrossEntropy as a loss function is motivated by the Neyman–Pearson lemma [88] that states that the best test statistics to distinguish two distributions is the log-likelihood ratio. The cross-entropy can be derived from that very ratio.

Backpropagation

The objective of the training is to find a set of DNN parameters that minimises the loss function. Theoretically, the global minimum is accessible analytically; however, this approach might be computationally not feasible. Backpropagation offers an alternative approach to determining a local minimum of the loss function. When using a DNN for predicting the class accordance of an object, data flows from the input layer through the network to the output layer. During the training, however, the data may flow backwards through the network for computing gradients; this procedure is called backpropagation [89]. DNNs are nested functions since the loss depends on the output layer, which depends on the previous layer, etc. Therefore, calculating gradients during backpropagation heavily depends on the chain rule. The gradient of the loss function is computed with respect to each model parameter θ_i .

$$\nabla_{\theta} L_{\{x,y\}}(\theta) = \begin{bmatrix} \frac{\partial L}{\partial \theta_0} \\ \vdots \\ \frac{\partial L}{\partial \theta_n} \end{bmatrix}$$

Then, all parameters of the DNN are updated with the negative gradient of the loss function multiplied by a step width, referred to as the learning rate α . Stepping in the direction of the negative gradient of the loss function minimises the loss function and consequently improves the prediction of the network, as depicted in Figure 4.5.

$$\theta' = \theta - \alpha \nabla_{\theta} L(\theta) \quad (4.2)$$

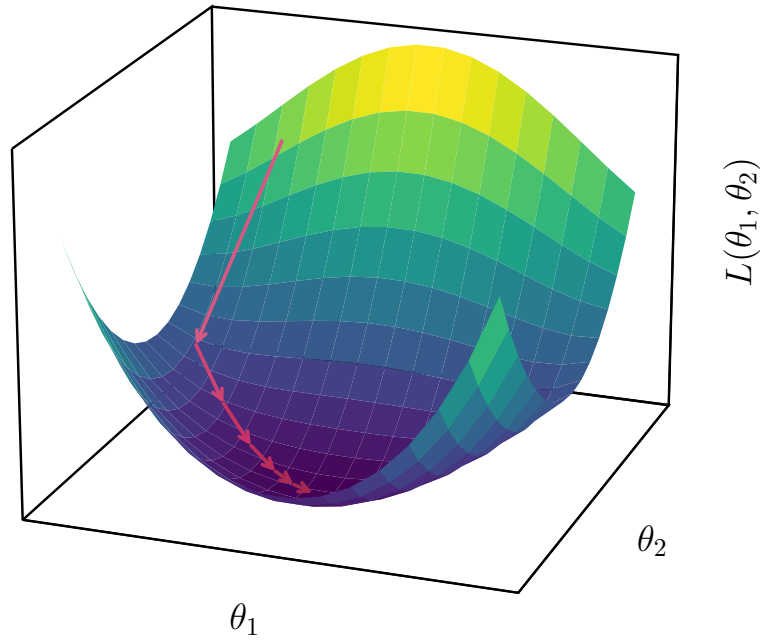


Figure 4.5.: Illustration of the loss value in dependence of the model parameters θ_1 and θ_2 . During training, the network parameters θ converge to a set of values, which minimises the loss function L . Each arrow indicates one optimisation step of the DNN's parameters in the direction of the negative gradient as defined in Equation 4.2.

The learning rate is a crucial hyperparameter that needs to be tuned during the development of a DNN. A too-small learning rate might slow down the network's training or lead to no update of the network's parameters. A too-high learning rate, however, might lead to divergent behaviour since the weight updates might overshoot the minimum of the loss function.

Usually, instead of computing the gradient of the DNN's parameters for the entire training sample at once, only a subset, the so-called batch, is used. The batch-wise approach requires less memory, and more training steps are possible for the same computational effort. However, the variance increases with decreasing batch size. Therefore, the optimal batch size has to be determined manually during the development of a DNN. The iteration over all batches (i.e. the total sample) is called an epoch. Multiple epochs may be required to converge to a set of model parameters that minimise the loss function.

4.4. Regularisation

The DNN's complexity (i.e. the number of nodes and hidden layers) is a crucial design choice. If the architecture is too small, the DNN might not be capable of learning the dependencies needed for assigning the samples to the corresponding class. On the other hand, large architecture may lead to overfitting. Overfitting refers to the problem of learning dependencies of the training sample that are only valid for that very sample but do not generalise to unseen data. Therefore, the dataset is split into a training and validation sample. Only the training sample is used to update the parameters via computing gradients. After each epoch, the loss on the training and the validation sample is compared. If the loss value on the training is lower than for the validation sample, the DNN has learned patterns that do not generalise to unseen data. The dependence of the generalisation gap between the training and validation sample is shown in Figure 4.6.

Regularisation summarises all techniques counteracting the overfitting of the training sample, including *Dropout* [90]. In Dropout, a predefined percentage of the nodes do not contribute to the classification task (i.e. all weights connected to that neuron are set to zero) while processing one batch. Using only a subset of the nodes for classification reduces the model complexity during training, which may reduce overfitting. Furthermore, using different nodes during the training can be considered a combination of independent models. Further regularisation techniques are summarized in Reference [91].

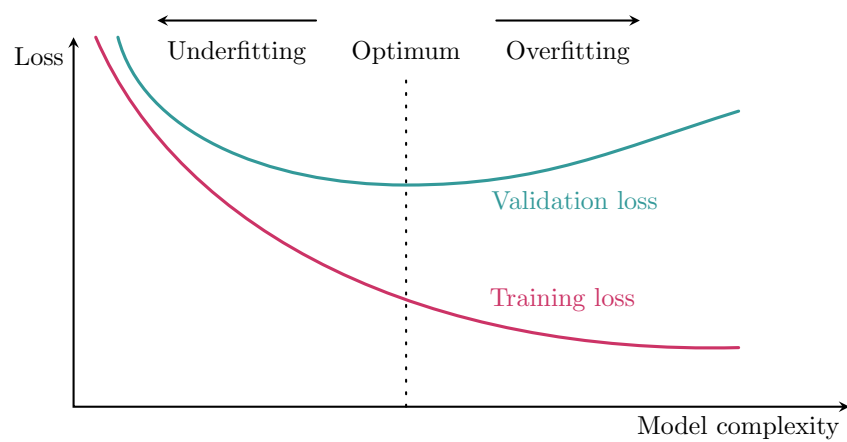


Figure 4.6.: Visualisation of the loss value in dependence on the model's complexity of a trained model. A small model might not be capable of expressing the dependencies in data. A very large model, however, might learn patterns that are only valid for the training sample (what is called overfitting). The best generalising network has the lowest loss value on the validation sample.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

The standard approach of reconstructing and identifying $Z \rightarrow e^+e^-$ decays suffers from significant deficiencies for boosted topologies. In a novel approach, a $Z \rightarrow e^+e^-$ candidate decay is reconstructed via a single jet, clustered via the anti- k_t algorithm using a radius parameter of 0.4 as depicted in Figure 5.1. For the identification, a DNN is trained, which operates on small-radius jet properties as well as inner detector and calorimeter information. The samples used to train the DNN are simulated with the Monte-Carlo event generator `SHERPA 2.2.11` [92]. The DNN has four-dimensional output corresponding the signal ($Z \rightarrow e^+e^-$) and each of the three background classes ($q/g, e/\gamma, (\tau)\tau$). Training with multiple output nodes offers high flexibility when constructing the final discriminant of the identification algorithm. Combining the probabilities of the signal and background classes allows for further optimising of the algorithm's performance.

At first, the limitations of the standard approach are detailed in Section 5.1. After that,

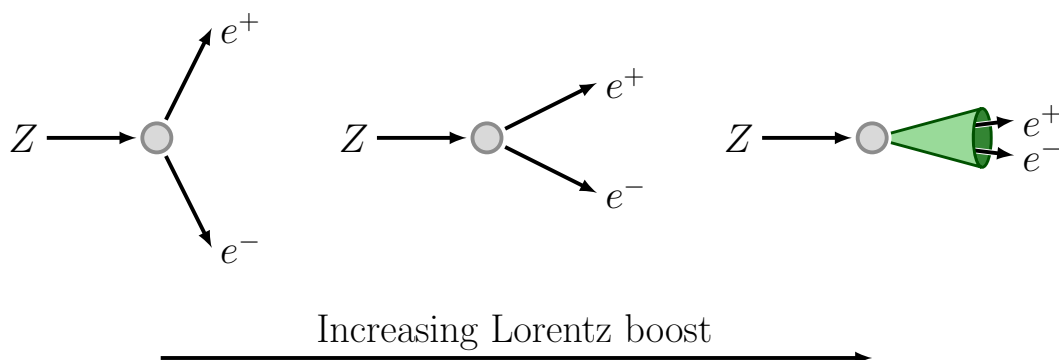


Figure 5.1.: A pictorial representation of $Z \rightarrow e^+e^-$ decays for different Lorentz boosts. A high Lorentz boost of the Z boson leads to a low angular separation between the e^+e^- pair. The green cone illustrates the small- R jet used to reconstruct the $Z \rightarrow e^+e^-$ decay.

the reconstruction of the $Z \rightarrow e^+e^-$ decay and the relevant backgrounds is described in Section 5.3. In Section 5.4 the input variables of the DNN are presented. The architecture and training of the DNN is outlined in Section 5.5. Section 5.6 concludes the signal efficiencies and background rejection rates as a function of jet properties. Finally, the relative importance of the input variables to separate signal and background is provided in Section 5.7.

5.1. Shortcomings of the Standard $Z \rightarrow e^+e^-$ Identification and Reconstruction

As outlined in Section 3.3.1, the ATLAS standard techniques for electron reconstruction are based on matching the track of a charged particle to energy clusters in the calorimeter system. The electron identification is based on a likelihood procedure that combines information on electromagnetic shower shapes, track quality, and track-cluster matching in order to separate prompt electrons from relevant backgrounds.

Both reconstruction and identification procedures are highly performant over a large

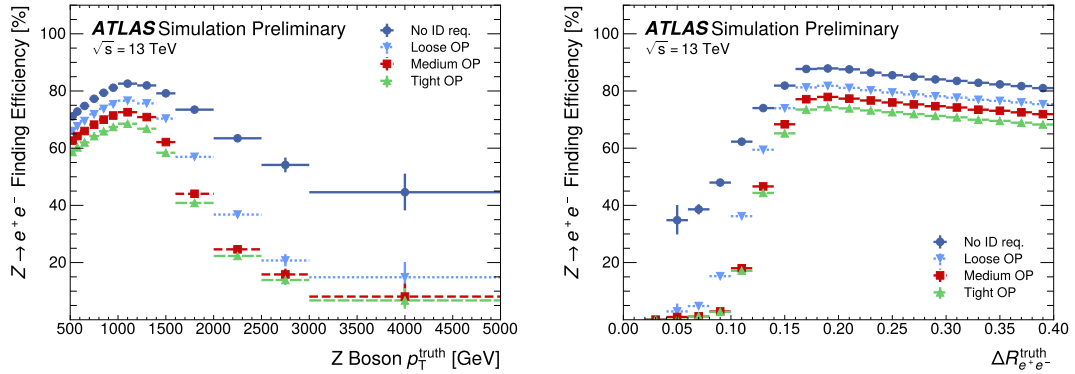


Figure 5.2.: Efficiency to reconstruct and identify $Z \rightarrow e^+e^-$ decays using standard ATLAS electron reconstruction and identification algorithms as a function of the generator-level Z boson p_T (right) and the angular separation $\Delta R(e^+, e^-)$ between the generator-level objects (right). The efficiencies are presented for the *Tight*, *Medium*, and *Loose* OPs and without applying any identification requirements. Electrons with a transverse energy larger than 25 GeV and a pseudorapidity satisfying $|\eta| < 2.47$, excluding the transition regions between calorimeters $|\eta| \in [1.37, 1.52]$, are considered.

kinematic range. However, the electron reconstruction and identification performance in highly boosted $Z \rightarrow e^+e^-$ decays is significantly limited, as shown in Figure 5.2. As the Lorentz boost of the Z boson increases and therefore the angular separation decreases, the efficiency to reconstruct both final state particles degrades dramatically. The efficiencies for identifying a prompt electron with $E_T = 40$ GeV are 93 %, 88 %, and 80 % for the Loose, Medium, and Tight operating points, respectively.

The degradation effects for large Z boson momenta (i.e. for strongly collimated e^+e^- systems) are driven by, e.g. the merging of energy clusters and distortion of shower shapes and track quality observables. As defined in Equation 3.2, the log-transformed LH discriminant combines the information of all relevant observables in a discriminant. Thus, changes in the individual observables are passed on to the discriminant and deviations are amplified. The log-transformed LH discriminant of electrons from $Z \rightarrow e^+e^-$ decays are separately shown in Figure 5.3. The blue histogram represents the log-transformed LH value of each electron from $Z \rightarrow e^+e^-$ decays with an angular separation between the electrons of $\Delta R < 0.1$, and the red histogram of each electron from $Z \rightarrow e^+e^-$ decays with an angular separation between the electrons of $\Delta R > 0.1$.

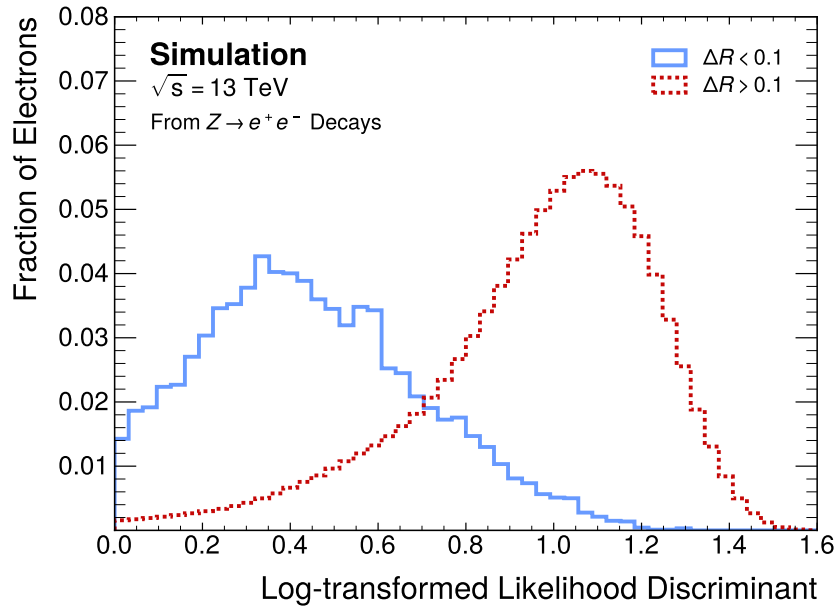


Figure 5.3.: Log-transformed likelihood discriminant of the reconstructed electrons from $Z \rightarrow e^+e^-$ decays with a separation between the two electrons of $\Delta R < 0.1$ (blue) and $\Delta R > 0.1$ (red).

The log-transformed LH discriminant distribution is shifted to lower values for a low angular separation between the electrons.

Various BSM theories predict the existence of new heavy vector bosons with significant branching fractions to the H , W and Z bosons. However, resonance searches are typically not sensitive to all free parameters of the underlying theory. Therefore, simplified model models, such as the heavy vector triplet (HVT) model (see Section 2.3.1), are used. The HVT framework adds an additional $SU(2)$ field to the SM with an electrically charged W' boson and an electrically neutral Z' boson. Several recent diboson resonance searches (such as searches for Z' and W' bosons) are based on channels targeting leptonic Z boson decays and were found to be limited by the inefficiency in reconstructing and identifying highly boosted $Z \rightarrow e^+e^-$ decays [19–21].

Figure 5.4, Figure 5.5, Figure 5.6 depict the product of acceptance and efficiency for $Z' \rightarrow Zh \rightarrow l^+l^-b\bar{b}/c\bar{c}$, $X \rightarrow ZV \rightarrow llq\bar{q}$ and $W' \rightarrow l\nu ll$ respectively. For masses at the TeV scale, a degradation of the product of acceptance and efficiency can be seen for all decay channels including e^+e^- pairs.

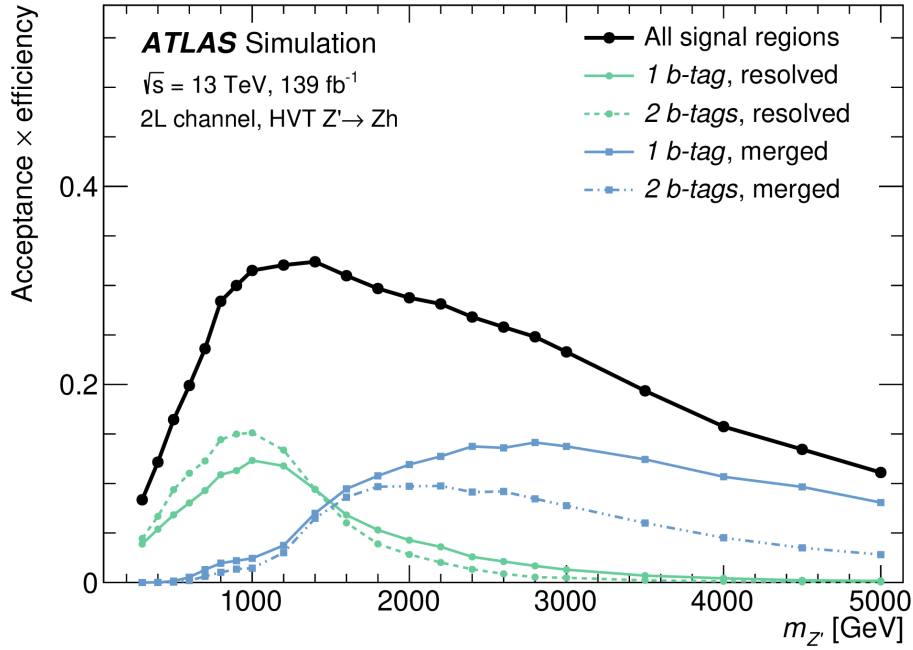


Figure 5.4.: Product of acceptance and efficiency as a function of the pole mass of an exotic heavy vector boson Z' for the $Z' \rightarrow Zh \rightarrow l^+l^-b\bar{b}/c\bar{c}$ decay channel. [20]

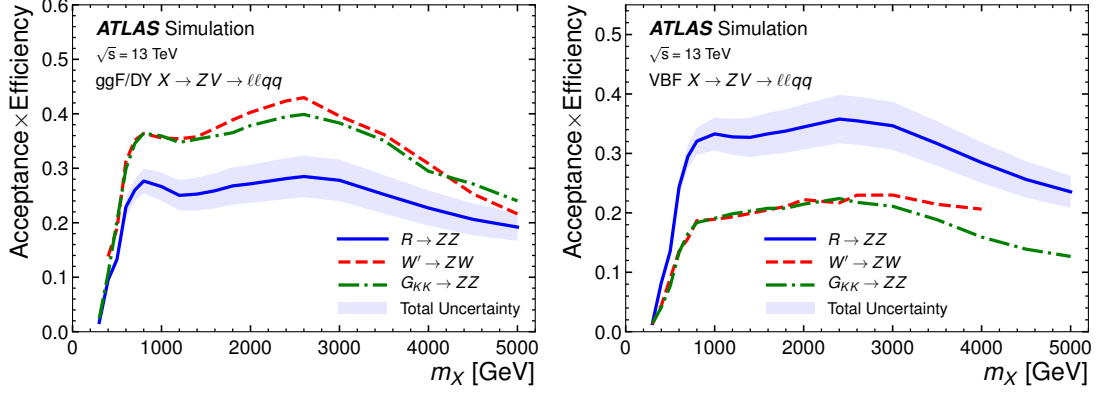


Figure 5.5.: Selection acceptance times efficiency for the $X \rightarrow ZV \rightarrow \ell\ell qq$ signal events from MC simulations as a function of the resonance mass for ggF/DY (right) and VBF production (left). [19]

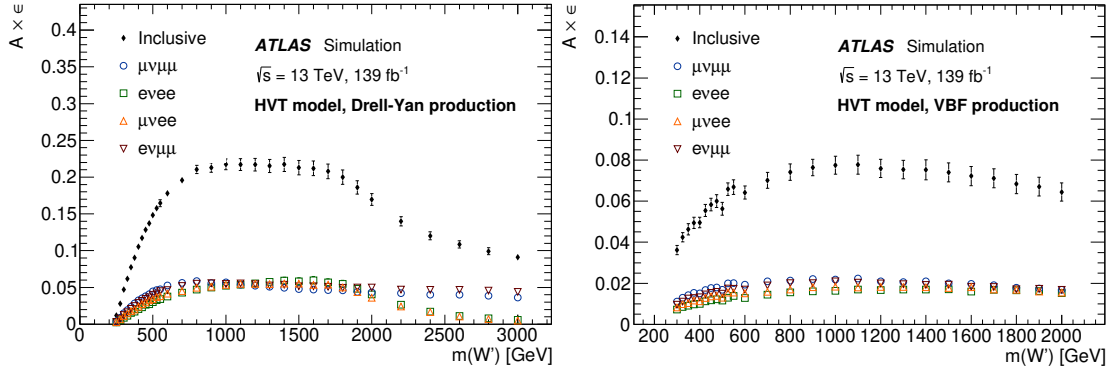


Figure 5.6.: The acceptance (A) times efficiency (ϵ) of the HVT W' in the Drell-Yan (left) and VBF (right) signal region for different mass points and for the individual channels, and the sum of all channels. The uncertainty includes both the statistical and experimental systematic components. [21]

5.2. The Monte Carlo Dataset

The development of the $Z \rightarrow e^+e^-$ tagging algorithm and the evaluation of its performance is based on the Monte Carlo (MC) simulation of various processes including the production of W and Z/γ^* bosons in association with jets or photons. All relevant MC samples are generated using SHERPA 2.2.11 [92] with the NNPDF3.0_{NNLO} PDF set. For W + jets and Z/γ^* + jets production, matrix elements are calculated at next-to-leading (NLO) precision in QCD, for diagrams with up to two additional parton emissions, and leading-order (LO) precision in QCD for diagrams with three, four or five additional parton emissions. For $W\gamma$ and $Z\gamma$ production, matrix elements are calculated at NLO precision in QCD for diagrams with up to one additional parton emission, and LO precision in QCD for diagrams with two, three, or four additional parton emissions. For both processes, Z boson decays to e^+e^- and $\tau^+\tau^-$, and W boson decays to $e^\pm\nu$ and $\tau^\pm\nu$ are considered. The matrix element calculation is based on the Comix [93] and OPENLOOPS [94–96] libraries, while the MEPS NLO prescription [97–100] is used to merge the matrix elements and the SHERPA parton shower [101], which is based on a set of tuned parameters developed by the SHERPA authors. All events are generated assuming proton-proton (pp) collisions at a centre-of-mass energy $\sqrt{s}$ of 13 TeV and passed afterwards through the full ATLAS detector simulation based on GEANT4 [102, 103]. The simulated events are overlaid with additional inelastic pp interactions in the same and neighbouring bunch crossings (pile-up). These interactions were simulated with the single-, double- and non-diffractive pp processes of PYTHIA8.186 using the A3 set of tuned parameters [104] and the NNPDF2.3_{LO} PDF. An overview of the MC samples used for the development of the DNN is given in Table 5.1.

Table 5.1.: The Monte Carlo samples used for the deployment of the $Z \rightarrow e^+e^-$ identification algorithm.

Process	Generator
$Z \rightarrow e^+e^-$	SHERPA2.2.11
$Z \rightarrow \tau^+\tau^-$	SHERPA2.2.11
$W^\pm \rightarrow e^\pm\nu$	SHERPA2.2.11
$W^\pm \rightarrow \tau^\pm\nu$	SHERPA2.2.11
$Z(\rightarrow l^+l^-)\gamma$	SHERPA2.2.11
$W^\pm(\rightarrow l\nu)\gamma$	SHERPA2.2.11

5.3. Object Reconstruction

The $Z \rightarrow e^+e^-$ identification algorithm separates signal from the underlying backgrounds based on jet properties, including ID and calorimeter information. This Section outlines the object reconstruction used for the identification algorithm as well as the preselection and truth labelling.

Jet Energy Resolution Studies Reconstructing $Z \rightarrow e^+e^-$ decays as jets is a so far uncommon approach. Therefore, the jet energy resolutions (JER) of four different cases are studied:

- Particle-flow jets calibrated to the EM scale including JES (EMPFLOW+JES)
- Uncorrected particle-flow jets calibrated to the EM scale (EMPFLOW)
- Uncorrected topo jets calibrated to the EM scale (EMTopo)
- Uncorrected topo jets calibrated to the LC scale (LCTopo)

Uncorrected refers to prior to applying pile-up corrections and the absolute MC-based corrections, which is a part of the JES corrections [105]. An overview of the jet calibration and correction techniques at ATLAS is provided in Section 3.3.4. All jets are clustered using the anti- k_t algorithm [74] as implemented in the FASTJET package [106] with a radius parameter of 0.4 [72]. The efficiency to reconstruct a $Z \rightarrow e^+e^-$ decay via a single anti- k_t -jet is depicted in Figure 5.7 as a function of the Z boson p_T . The efficiencies range from 30 % for a Z boson with p_T of 500 GeV up to 90% for a Z boson with p_T of 3000 GeV.

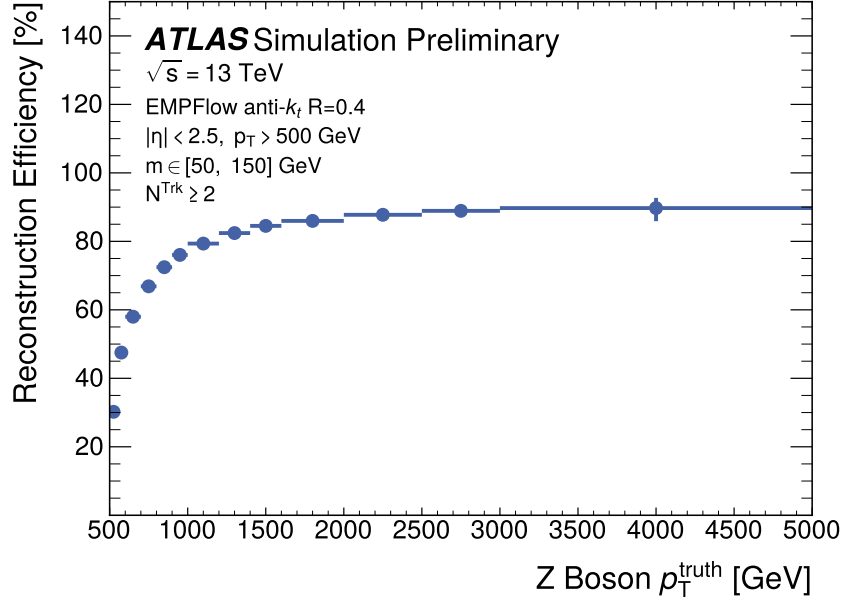


Figure 5.7.: Efficiency to reconstruct a $Z \rightarrow e^+e^-$ decay via a single anti- k_t -jet with $p_T > 500 \text{ GeV}$, $|\eta| < 2.5$, $N^{\text{Trk}} \geq 2$, and $50 \text{ GeV} < m < 150 \text{ GeV}$ presented as a function of the generator-level Z boson p_T . The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

The jets were studied in terms of the mass and transverse momentum response [107]:

$$\mathcal{R}_m = \frac{m^{\text{reco}}}{m^{\text{truth}}} \qquad \mathcal{R}_{p_T} = \frac{p_T^{\text{reco}}}{p_T^{\text{truth}}}$$

The masses and transverse momenta at generator-level, m^{truth} and p_T^{truth} , are calculated via all generator-level particles ghost-associated [108, 109] with the reconstructed jets. Jets consist of pointlike particles, which themselves have no intrinsic area. Therefore, the sensible area is defined by adding infinitely soft particles (the so-called ghosts) and identifying the region in η , ϕ where those ghosts are clustered with a given jet. The extent of this region provides a measure of the jet area.

The mass and momentum response scales (μ) and resolutions (σ) are determined by fitting the convolution of a Gaussian and an Exponential function (referred to as Bukin

function [110]) through these mass and momentum response distributions.

The EMPflow+JES calibration is the standard jet calibration at ATLAS for reconstructing EM objects. Figure 5.8 and Figure 5.9 show the mass and momentum response distribution respectively for $Z \rightarrow e^+e^-$ -jets with $p_T > 1000$ in four $|\eta|$ intervals. In the transition regions ($|\eta| \in [1.37, 1.52]$), the momentum response distribution shows two local maxima, and in the forward regions ($|\eta| \in [1.52, 2.5]$), μ is shifted to values of ~ 1.1 . The JES can explain these effects since the pile-up corrections and the absolute MC-based corrections are η , p_T -dependent and not optimised for $Z \rightarrow e^+e^-$ -jets.

Figure 5.10 and Figure 5.11 show the mass and momentum response distributions for each of the three uncalibrated jets (EMPflow, EMTopo, and LCTopo) in four different $|\eta|$ intervals for $p_T \in [1000, 2500]$ GeV. Figure 5.12, Figure 5.13, Figure 5.14 depict the response scales μ with resolutions σ as errorbars for each of the three cases in twelve η , p_T intervals. Most noticeable are the different response scales of the EM and LC calibration. For jets calibrated to the EM scale, μ has values < 1 , while jets calibrated to the LC scale have values $\mu > 1$. However, the absolute value of the biases is similar ($\sim 3\%$). Hadronic objects deposit energy via the strong interaction in the detector. Energy deposited via the strong interaction does not result in a detector signal. Hence, the response of uncalibrated hadronic jets is lower than for electromagnetic jets. To compensate for this effect, the magnitudes of the LC calibration are higher than for the EM calibration. The LC calibration applied to the EM objects ($Z \rightarrow e^+e^-$ -jets) leads to an overestimation of the energy (see Figure 5.12). The underestimation of the energy of the EM jets indicates hadronic activity in the leptonic $Z \rightarrow e^+e^-$ -jets.

Except in the transition regions ($|\eta| \in [1.37, 1.52]$), the EM calibration has a better resolution than LC. In the transition regions, however, the LC calibration profits from the dead material corrections which are not included in the calibrations to the EM scale.

EMPflow has a slightly higher resolution (i.e. lower σ) than EMTopo due to the combination of the advantages of the ID and calorimeter.

Therefore, the *uncorrected* EMPflow jets are used to reconstruct $Z \rightarrow e^+e^-$ decays.

As shown in Figure 5.14, the mass and momentum response scales of the EMPflow jets are close to unity for most of the probed kinematic range, except for the transition regions between calorimeters ($|\eta| \in [1.37, 1.52]$), where the response scales are at $\sim 95\%$. Both the momentum and mass resolution vary mildly as a function of the jet p_T and $|\eta|$. For $|\eta| < 1.37$, the mass (momentum) resolution varies between 0.7% and 1.1% (2.5% and 3.8%), while for $|\eta| > 1.52$ the momentum (mass) resolution varies 1.4% and 1.5% (3.2% and 4.9%). Larger variations are found in the transition regions, where the momentum (mass) resolution varies between 4.4% and 5.2% (4.6% and 6.1%).

Figure 5.15 shows a mild dependence of the mass and p_T scale on the number of interaction vertices; the resolution is stable as a function N^{Vert} .

The good overall agreement between the reconstructed and expected jet momenta and masses justifies the use of uncorrected EMPflow jets for the development of the $Z \rightarrow e^+e^-$ identification algorithm. However, dedicated pile-up and MC-based corrections will have to be applied once the algorithm is used in physics analyses.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

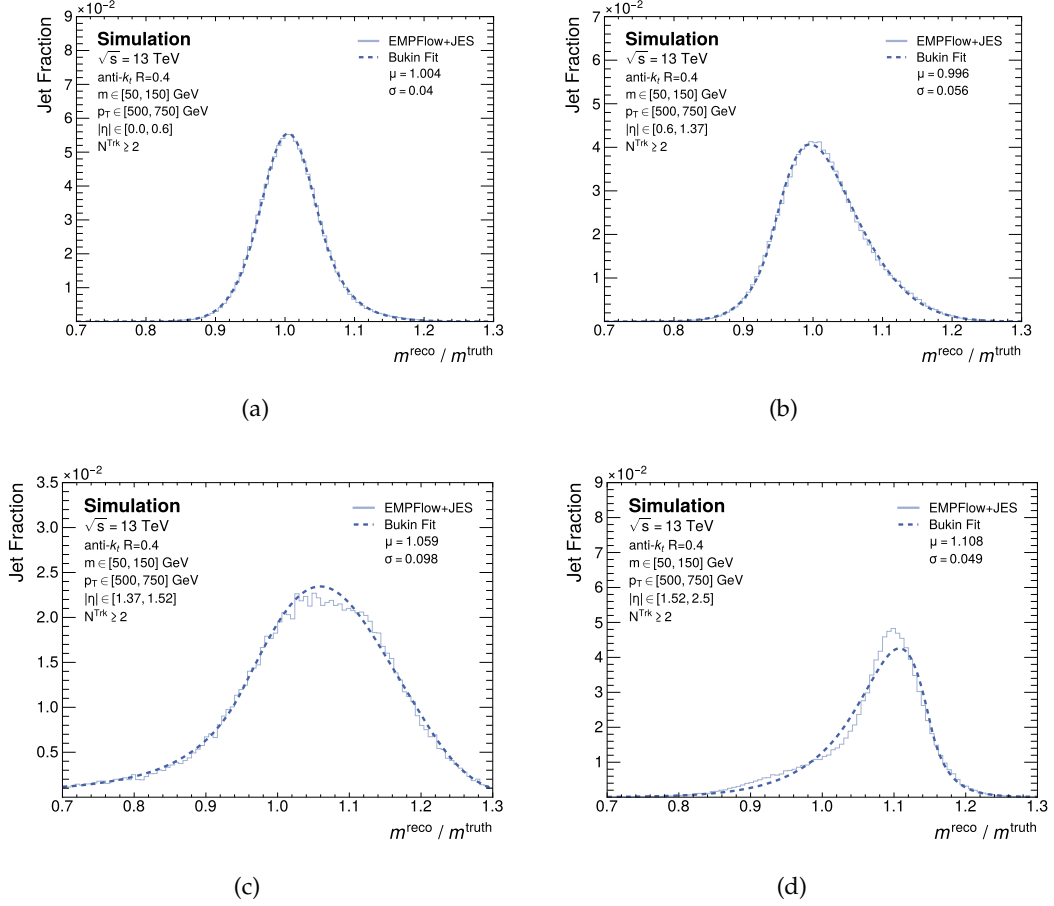


Figure 5.8.: Distributions of $m^{\text{reco}} / m^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on particle-flow objects calibrated to the EM scale including JES, with $p_T \in [500, 750]$ GeV and (a) $|\eta| \in [0.0, 0.6]$, (b) $|\eta| \in [0.6, 1.37]$, (c) $|\eta| \in [1.37, 1.52]$ and (d) $|\eta| \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0_{NNLO} PDF set. The mass response distributions of further intervals are presented in Figure A.4 in the Appendix.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

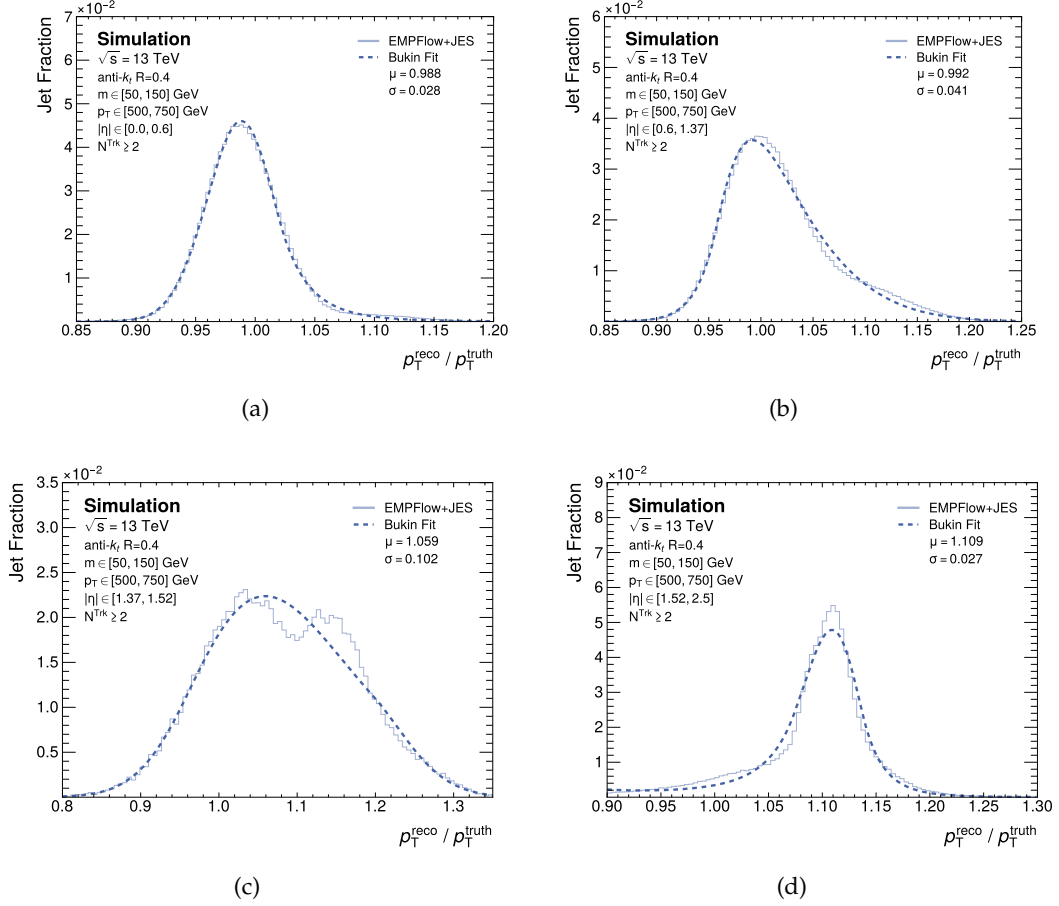


Figure 5.9.: Distributions of $p_T^{\text{reco}} / p_T^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on particle-flow objects calibrated to the EM scale including JES, with $p_T \in [500, 750]$ GeV and (a) $|\eta| \in [0.0, 0.6]$, (b) $|\eta| \in [0.6, 1.37]$, (c) $|\eta| \in [1.37, 1.52]$ and (d) $|\eta| \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The momentum response distributions of further intervals are presented in Figure A.5 in the Appendix.

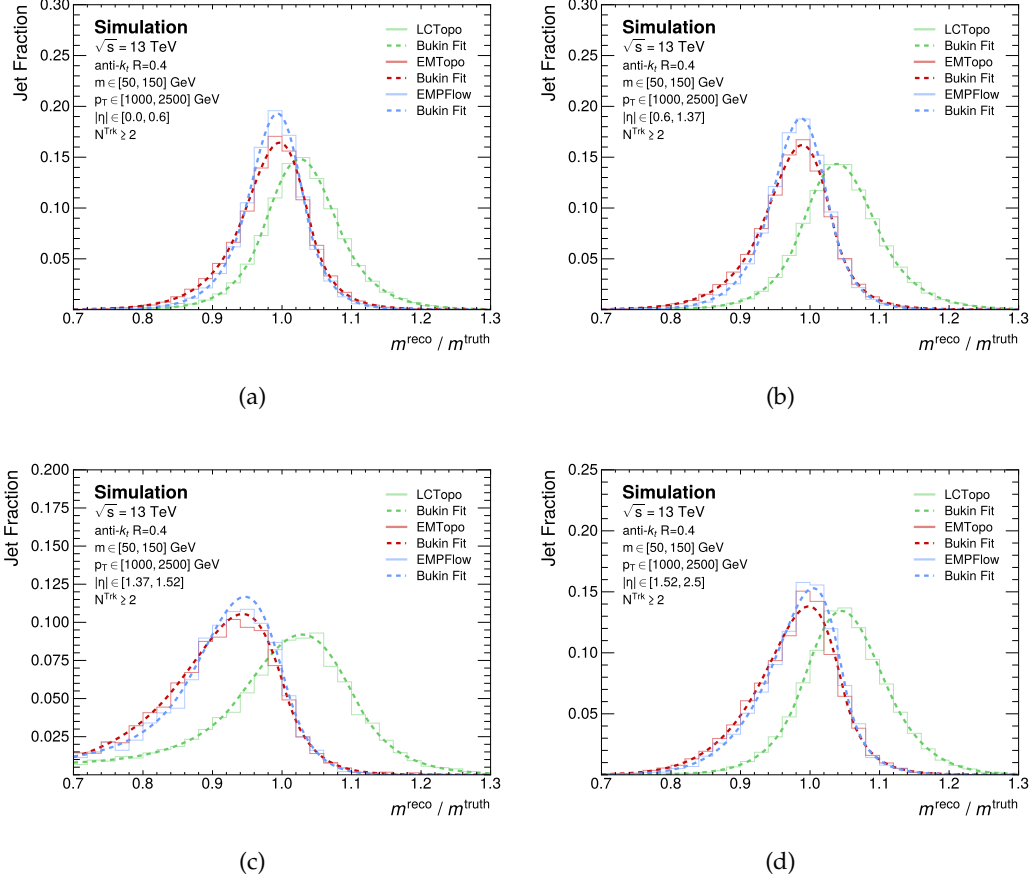


Figure 5.10.: Distributions of $m^{\text{reco}} / m^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on uncorrected topo jets calibrated to the LC scale (green), uncorrected topo jets calibrated to the EM scale (red) and uncorrected particle-flow jets calibrated to the EM scale (blue). Each case is depicted for $p_T \in [1000, 2500]$ GeV and (a) $|\eta| \in [0.0, 0.6]$, (b) $|\eta| \in [0.6, 1.37]$, (c) $|\eta| \in [1.37, 1.52]$ and (d) $|\eta| \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The momentum response distributions of further intervals are presented in Figure A.6 in the Appendix.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

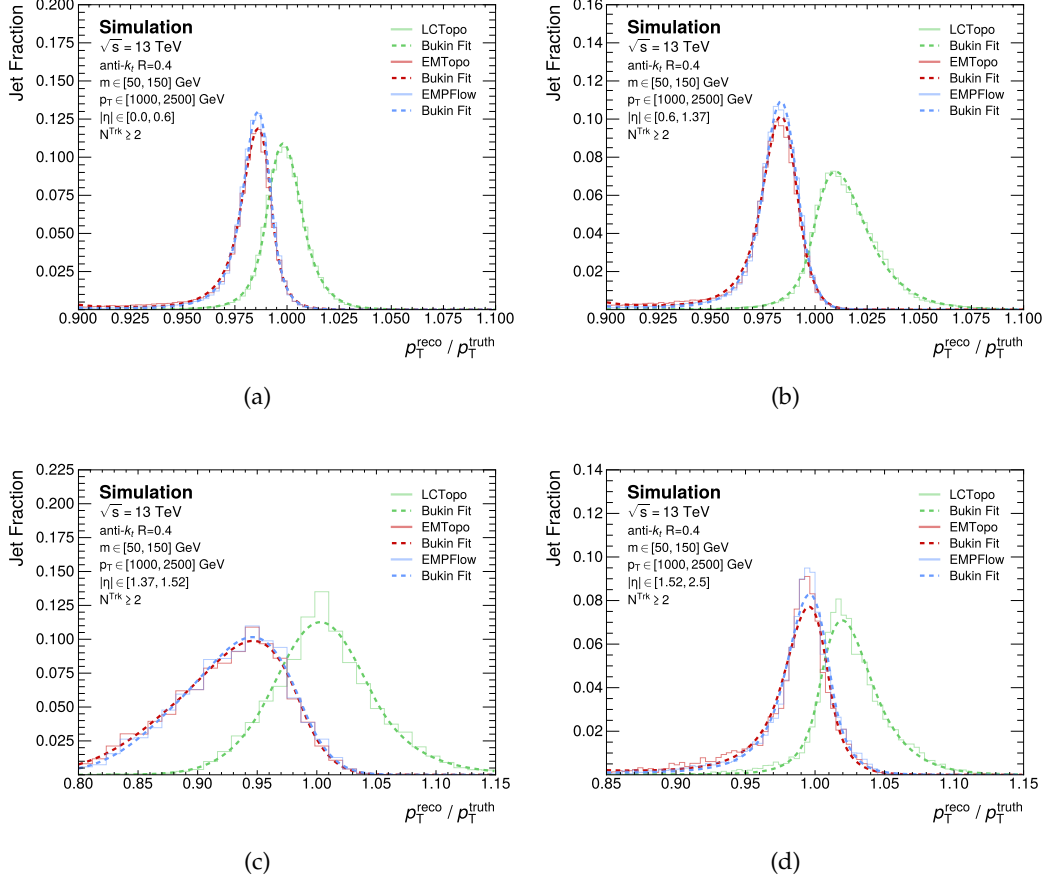


Figure 5.11.: Distributions of $p_T^{\text{reco}} / p_T^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on uncorrected topo jets calibrated to the LC scale (green), uncorrected topo jets calibrated to the EM scale (red) and uncorrected particle-flow jets calibrated to the EM scale (blue). Each jet type is depicted for $p_T \in [1000, 2500]$ GeV and (a) $|\eta| \in [0.0, 0.6]$, (b) $|\eta| \in [0.6, 1.37]$, (c) $|\eta| \in [1.37, 1.52]$ and (d) $|\eta| \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The momentum response distributions of further intervals are presented in Figure A.7 in the Appendix.

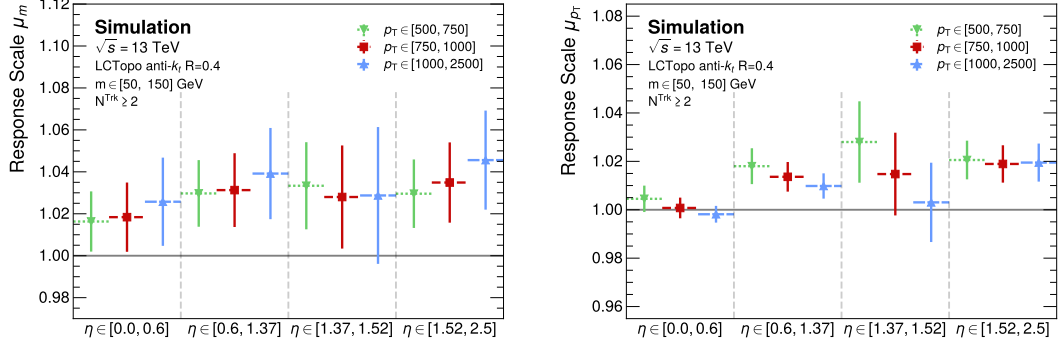


Figure 5.12.: Response scale μ of the twelve η , p_T intervals of the uncorrected topo jets calibrated to the LC scale of the mass (left) and transverse momentum (right). The error bars show the resolutions σ . The underlying response distributions are depicted in Figure A.6 and Figure A.7 in the Appendix. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0_{NNLO} PDF set.

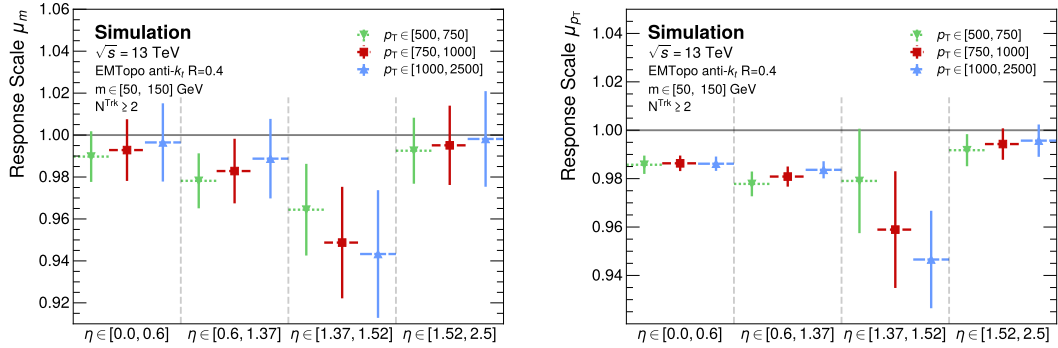


Figure 5.13.: Response scale μ of the twelve η , p_T intervals of the uncorrected topo jets calibrated to the EM scale of the mass (left) and transverse momentum (right). The error bars show the resolutions σ . The underlying response distributions are depicted in Figure A.6 and Figure A.7 in the Appendix. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0_{NNLO} PDF set.

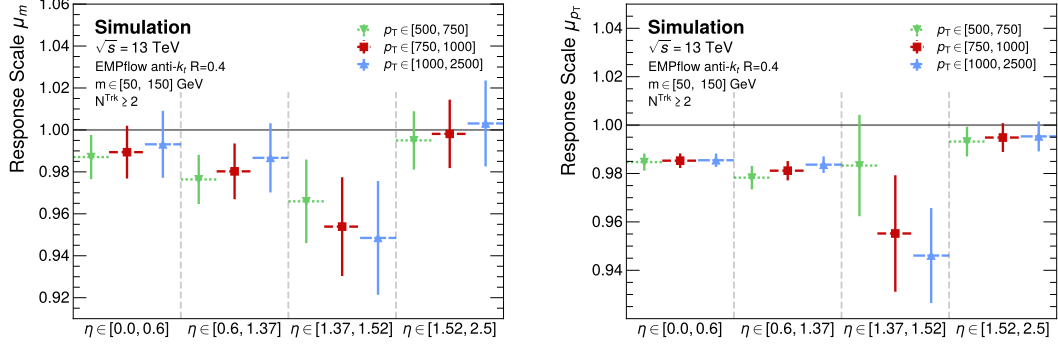


Figure 5.14.: Response scale μ of the twelve η , p_T intervals of the uncorrected particle-flow jets calibrated to the EM scale of the mass (left) and transverse momentum (right). The error bars show the resolutions σ . The underlying response distributions are depicted in Figure A.6 and Figure A.7 in the Appendix. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

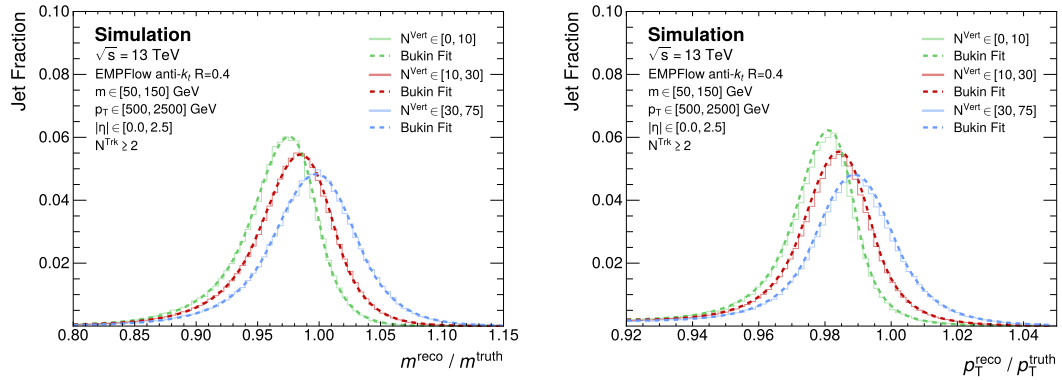


Figure 5.15.: Mass (left) and transverse momentum response distributions (right) for $N^{\text{Vert}} \in [0, 10]$ (red), $N^{\text{Vert}} \in [10, 20]$ (green) and $N^{\text{Vert}} \in [20, 75]$ (blue). The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

Superclusters As for the standard electron reconstruction, variable-sized superclusters [111] are formed dynamically from topologically connected cells in the calorimeter system [71], referred to as topo-clusters¹. The supercluster reconstruction procedure is based on a two-stage approach: in the first stage, a seed cluster is selected from a list of EM topo-clusters; in the second stage, EM topo-clusters near the seed are identified as satellite clusters, which may arise from bremsstrahlung radiation or splitting of topo-clusters. The final superclusters are obtained by adding the satellite clusters to the seed cluster if they satisfy the necessary selection criteria [111].

Tracks The reconstruction of charged-particle tracks in the ID is based on combining space-point measurements in the pixel, microstrip, and TRT detectors [65]. Track candidates with $p_T > 400$ MeV are fitted, according to the pion or electron hypothesis (depending on various quality requirements [31]) using the ATLAS Global χ^2 Track Fitter [66]. Tracks are required to have a minimum transverse momentum of 1 GeV, at least seven hits in the silicon detectors (of which one has to be in the most inner layer of the pixel detector), and to originate from the PV. The latter condition is satisfied by a requirement on the transverse impact parameter, $d_0 < 1.0$ mm, and on the longitudinal impact parameter, $|z_0 \cdot \sin(\theta)| < 1.5$ mm, where d_0 (z_0) represents the transverse (longitudinal) perigee parameter defined at the point of the closest approach of the trajectory to the z-axis [112].

Furthermore, tracks refitted with the refitting algorithm, based on a Gaussian sum filter (GSF), as used in the standard electron reconstruction (see Section 3.3.1), were tested. However, the GSF tracks were found to be unsuitable since the reconstruction efficiency is substantially degraded for low angular separation between the e^+e^- pair (see Figure A.1 in the Appendix).

Pre-selection The $Z \rightarrow e^+e^-$ candidates are reconstructed from a single anti- k_t -jet that is required to have $p_T > 500$ GeV, $|\eta| < 2.5$ and an invariant mass m within a window of $50 \text{ GeV} < m < 150 \text{ GeV}$. Furthermore, the jet must have at least two tracks within a cone of $\Delta R = 0.4$ around the jet axis, where the track momenta, defined at the perigee, are used for calculating the angular separation. Additionally, superclusters within a cone of $\Delta R = 0.4$ around the jet axis are exploited to calculate input quantities to the $Z \rightarrow e^+e^-$ identification algorithm.

¹Topo-clusters are three-dimensional, massless objects built via a nearest-neighbour algorithm using calorimeter cells as input. Cells are added to a topo-cluster if their energy content is significantly larger than the expected noise contribution [71].

Truth Labeling Labels are attributed to the jets in the simulation based on the following criteria. For the signal, jets are labeled as $Z \rightarrow e^+e^-$ -jets, if they contain exactly two prompt generator-level electrons within a cone of size $\Delta R = 0.4$ around the jet axis and with an invariant mass in the window $71\text{GeV} < m_{ee}^{\text{truth}} < 111\text{GeV}$. In addition, three types of backgrounds were identified to potentially mimic the signature of a $Z \rightarrow e^+e^-$ -jet. Jets containing a single prompt generator-level electron or photon within $\Delta R = 0.4$ to the jet axis are classified as e/γ -jets, while jets containing one or more prompt generator-level τ lepton within $\Delta R = 0.4$ to the jet axis are classified as $(\tau)\tau$ -jets². Jets that do not contain any prompt generator-level electron, photon, or τ -lepton are classified as quark- or gluon-jets (i.e. q/g -jets).

Figure 5.16 summarises the data set, the reconstruction, the pre-selection and the truth labelling used for the deployment of the new algorithm to identify highly boosted $Z \rightarrow e^+e^-$.

² $(\tau)\tau$ -jets in which both τ -leptons decay into an electron/positron are not considered in the following studies.

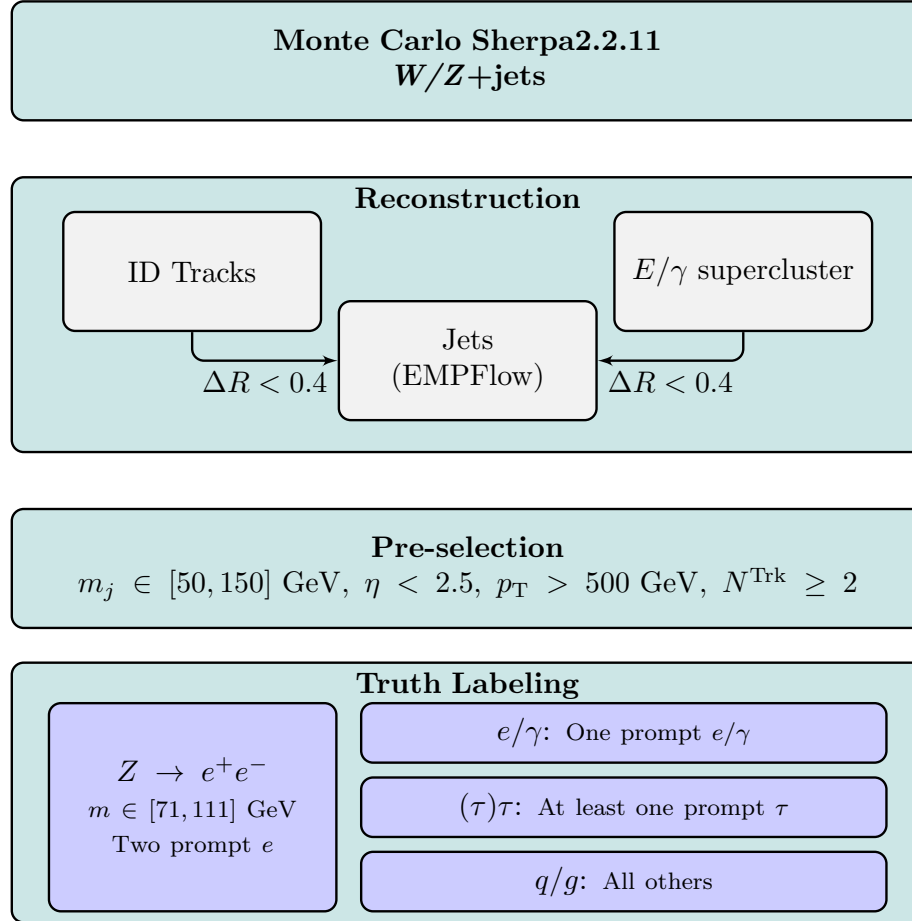


Figure 5.16.: The deployment of the DNN is based on Monte Carlo Samples generated with SHERPA 2.2.11. Jets, tracks and clusters are extracted from the MC sample and combined to one object via ΔR -matching. After the pre-selection of the $Z \rightarrow e^+e^-$ candidate jets, the jets are attributed to the four classes.

5.4. Input Variables

The DNN separates the signal jets ($Z \rightarrow e^+e^-$) and three background classes (q/g , e/γ , $(\tau)\tau$) based on the 29 observables as presented in the Table 5.2. The properties of the jet, the two leading tracks, and the superclusters associated with the jet are exploited. The set of track- and supercluster-based observables is inspired by the standard electron identification. Figures 5.18, 5.19, and 5.20 show the distributions of the input quantities separately for the signal class and the three background classes. The unique signature of $Z \rightarrow e^+e^-$ -jets will be discussed in the following.

Observables such as the number of tracks N^{Trk} and constituents $N^{\text{Const.}}$ illustrate the two-prong structure (signal jets peak at two). The angular separation between e^+e^- pair is approximately

$$\Delta R_{e^+e^-} \approx \frac{2m^Z}{p_T^Z}, \quad (5.1)$$

with the transverse momentum p_T^Z the mass m^Z of the Z boson [113, 114]. This property of the $Z \rightarrow e^+e^-$ decay is reflected in observables such as the angular separation between the two leading tracks $\Delta R(t_1, t_1)$, the angular separation between the jet centre and the lading (subleading) track $\Delta R(t_1, j)$ ($\Delta R(t_2, j)$) and the *Width*.

The *EM fraction* f_{EM} is defined as the energy deposited in the EM calorimeter normalised to the total jet energy. The EM fraction shows significant separation for the signal jets from the jets with hadronic activity (q/g and $(\tau)\tau$) as well as from the e/γ -jets. The electrons of the single electron jets carry the full p_T while the p_T of $Z \rightarrow e^+e^-$ decay is distributed over two electrons. Thus, single electrons deposit energy in the hadronic calorimeter as they penetrate deeper into the detector.

The correlation of the transverse momentum of the two electrons from the $Z \rightarrow e^+e^-$ decay leads to a distinct jet substructure encoded in unique symmetries. Observables sensitive to the two-prong jet substructure of the $Z \rightarrow e^+e^-$ -jets are for example the *ratio of the energy correlation functions* [115] and the *Planar Flow* [116, 117]; both quantities are determined based on the two leading tracks and are shown in Figure 5.18 and Figure 5.19

The energy correlation function for an arbitrary number of subjects N is defined as:

$$\text{ECF}(N, \beta) = \sum_{i_1 < i_2 < \dots < i_N \in J} \left(\prod_{a=1}^N p_{Ti_a} \right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^N R_{i_b i_c} \right)^\beta \quad (5.2)$$

with the angular distance between track i and j being R_{ij} , the angular exponent β and

the total number of particles J . When limited to the case of two tracks, the equation can be simplified to a more intuitive form:

$$r_{N=1}^{(\beta=1)} = \frac{\text{ECF}(2,1)}{\text{ECF}(1,1)} = \frac{\sum_{i < j \in J} p_{T_i} p_{T_j} R_{ij}}{\sum_{i \in J} p_{T_i}} = \frac{p_{T,1} p_{T,2}}{p_{T,1} + p_{T,2}} R_{12}, \quad (5.3)$$

For a fixed sum of the transverse momenta of the two leading tracks and a fixed separation R_{12} , this term is maximised for evenly distributed energy over the two leading tracks (i.e. $p_{T,1} = p_{T,2}$). Therefore, the distribution is shifted to higher values for the signal jets (see Figure 5.20).

The Planar Flow[116, 117] quantifies the degree to which the jet's energy is evenly spread over the area of the jet (values around 1) versus spread linearly across the area of the jet (values around 0). While the ratio of the energy correlation functions processes the transverse momenta of the tracks p_T , the Planar Flow is based only on the momentum components orthogonal to the jet axis z (not to be confused with the transverse momentum p_T , orthogonal to the beam axis), as depicted in Figure 5.17. At first, the event shape tensor I_w is defined as

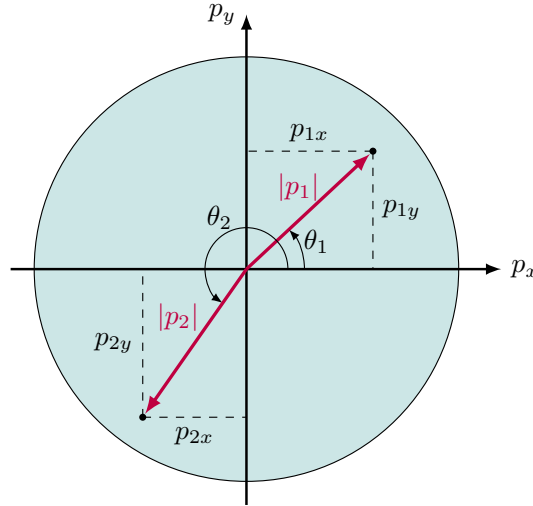


Figure 5.17.: Visualisation of the momenta transverse to the jet axis. The jet axis defines the origin of the coordinate system.

$$I_w^{kl} = \sum_i w_i \frac{p_{i,k}}{w_i} \frac{p_{i,l}}{w_i} \quad (5.4)$$

with $p_{i,k}$, the k^{th} momentum component of track i . Then, by multiplying I_w with $4/\text{tr}(I_w)^2$, the Planar Flow is accessible

$$Pf = \frac{4 \det(I_w)}{\text{tr}(I_w)^2}. \quad (5.5)$$

For the two leading tracks, Pf can be simplified as

$$Pf = \frac{|p_1|^2 E_2 \cdot |p_2|^2 E_1}{(|p_1|^2 E_2 + |p_2|^2 E_1)^2} \sin^2(\theta_2 - \theta_1). \quad (5.6)$$

The derivation is presented in Appendix A.3. For high angular separation between the two tracks ($\theta_2 - \theta_1 \rightarrow \pi$), the Planar Flow converges to zero. For the $Z \rightarrow e^+e^-$ -jets, the energy is mainly deposited by the electrons, leading to a relatively linear energy deposition ($Pf \rightarrow 0$). In contrast, the energy of the background jets is more planar distributed over the jet front ($Pf \rightarrow 1$).

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

Table 5.2.: Description of the quantities used to identify highly boosted $Z \rightarrow e^+e^-$ decays via a deep neural network. The track properties and the observables based on the track-cluster matching are considered for both the leading and subleading tracks per jet.

Input type	Observable(s)	Definition
Jet properties	m	Mass of the jet
	Width	Width of the jet defined as the p_T weighted average of the ΔR distances between the neutral particle flow objects and jet axis direction [105]
	p_T	Transverse momentum of the jet
	$ \eta $	Absolute pseudorapidity of the jet
	f_{EM}	Fraction of energy deposited in the EM calorimeter [118]
	$\max(E_{Layer}/E_{Jet})$	Maximum fraction of energy deposited in a single layer of the EM calorimeter [118]
	$N^{Const.}$	Multiplicity of particle-flow objects per jet
Track properties	d_0	Transverse impact parameter
	σ_{d_0}	Transverse impact parameter significance
	z_0	Longitudinal impact parameter
	σ_{z_0}	Longitudinal impact parameter significance
	p_e^{HT}	Probability that a track corresponds to an electron based on transition radiation in the TRT
	$\Delta R(t_i, j)$	Angular distance between the i -th leading track and the jet axis
Track-cluster matching	E/p	Ratio of the cluster energy to the track momentum
	$\Delta\Phi$	Azimuthal angular difference between the cluster position and the extrapolated track
Others	N^{Trk}	Multiplicity of tracks per jet
	Q^{Sum}	Sum of charges of the two leading tracks
	$\Delta R_{l_1 l_2}$	Angular separation between the two leading tracks
	Planar Flow	Planar Flow[116, 117] calculated from the two leading tracks. This observable quantifies the degree to which the jet's energy is evenly spread over the area of the jet (values around 1) versus spread linearly across the area of the jet (values around 0)
	$r_{N=1}^{(\beta=1)}$	Ratio of energy correlation functions $ECF(2,1)/ECF(1,1)$ [115] calculated from the two leading tracks. This observable is sensitive to the N-prong jet substructure
	Balance	$\frac{E_{Cluster}^{1st} - E_{Cluster}^{2nd}}{E_{Cluster}^{1st} + E_{Cluster}^{2nd}}$

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

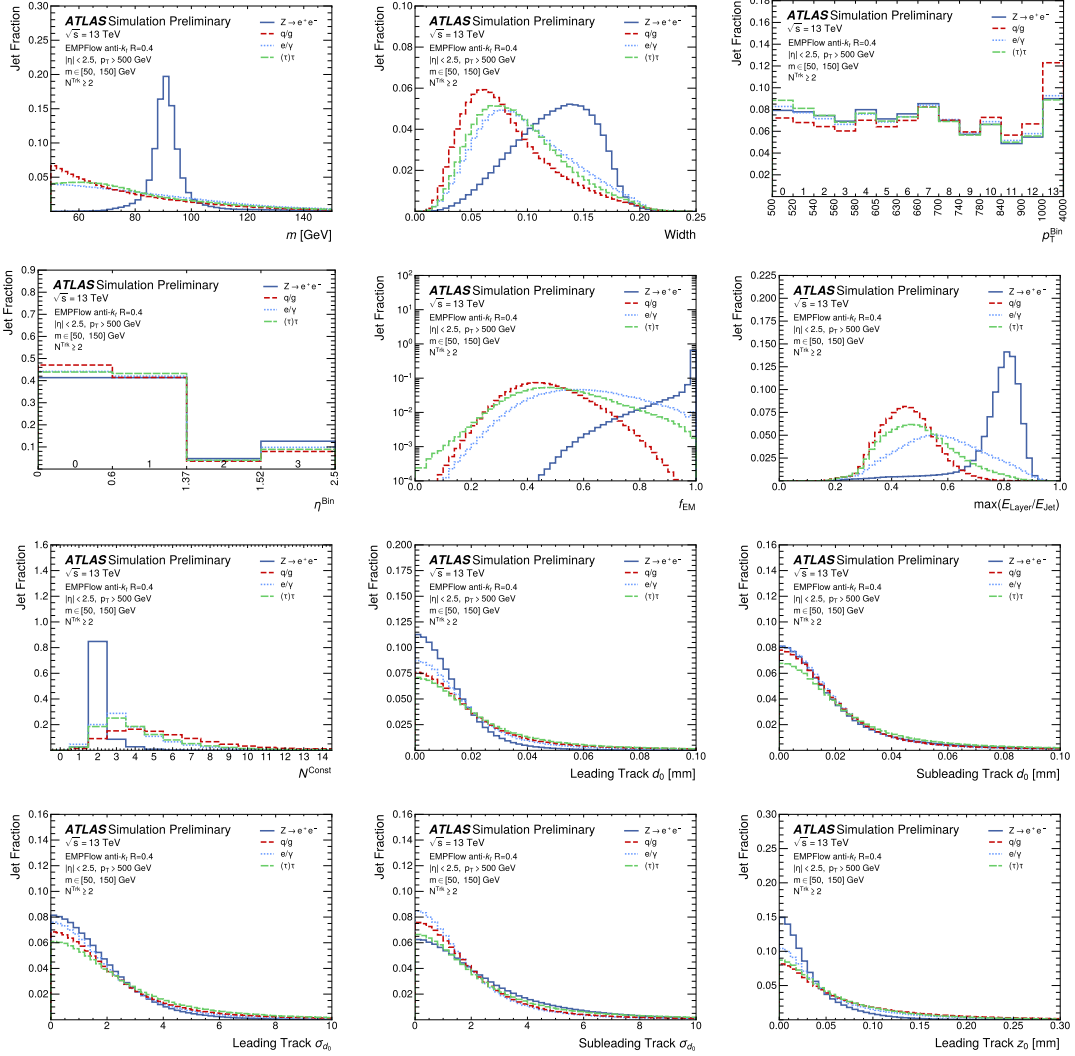


Figure 5.18.: Several inputs distributions as defined in Table 5.2. The distributions are shown before each jet type is individually reweighted to have flat p_T and $|\eta|$ spectra. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. All distributions are normalised to unity.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

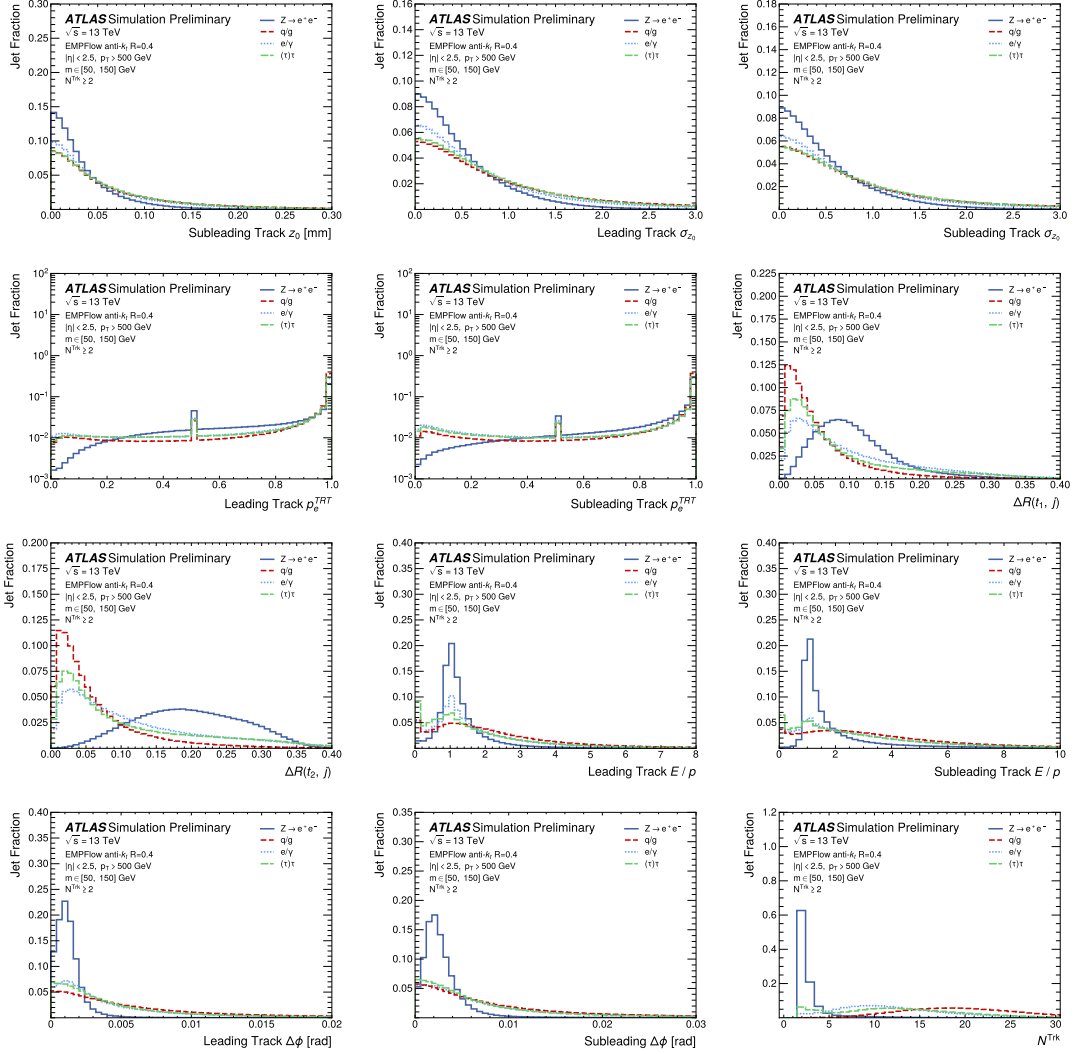


Figure 5.19.: Several inputs distributions as defined in Table 5.2. The distributions are shown before each jet type is individually reweighted to have flat p_T and $|\eta|$ spectra. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. All distributions are normalised to unity.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

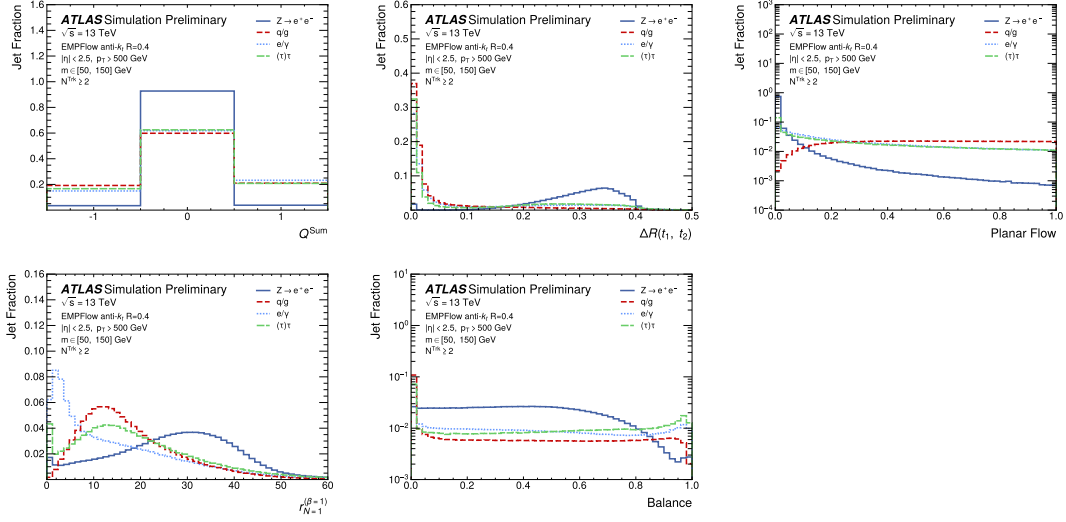


Figure 5.20.: Several inputs distributions as defined in Table 5.2. The distributions are shown before each jet type is individually reweighted to have flat p_T and $|\eta|$ spectra. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0_{NNLO} PDF set. All distributions are normalised to unity.

5.5. Network Architecture and Training Procedure

The $Z \rightarrow e^+e^-$ identification algorithm is based on a feed-forward neural network trained with the ADAM optimiser [119] and the CROSSENTROPYLoss function [120] using the machine learning framework PyTorch [121]. The DNN maps the input vector to a four-dimensional output layer via four hidden layers with 200 nodes each, as depicted in Figure 5.21. All hidden layers are batch normalised before the values of each node are fed into the SIGMOID activation function [122]. Furthermore, the SOFTMAX activation function [123] is used such that the sum of all outputs is equal to one. The four output nodes quantify the probabilities for a jet to correspond to the signal class and either of the three background classes. For numerical stability, the DNN output was log-transformed before applying the CROSSENTROPYLoss function.

The dataset used in the training consists of around 2 million $Z \rightarrow e^+e^-$ -jets, around 0.6 million $(\tau)\tau$ -jets, around 2 million e/γ -jets and around 4 million q/g -jets. The input dataset is randomly split into a training and a validation sample with a ratio of 6 : 1, which are compared to probe for signs of overtraining. The data used for the training

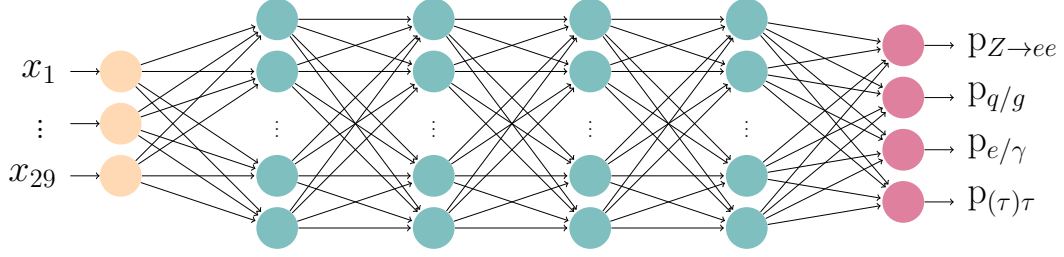


Figure 5.21.: Schematic of the DNN architecture containing a 29-dimensional input vector, four hidden layers with 200 nodes each and four output nodes. The output nodes quantify the probabilities for a jet to correspond to the signal ($Z \rightarrow e^+e^-$) or one of the three background classes (q/g , e/γ , $(\tau)\tau$).

is not used for the evaluation of the DNN. The p_T and $|\eta|$ are used as input variables of the DNN to take advantage of the correlations with the other observables. However, to prevent the network from learning differences in the kinematic distributions of the signal and background jets, each jet type is individually reweighted to have flat p_T^{Bin} and $|\eta^{\text{Bin}}|$ spectra. The interval boundaries in p_T and $|\eta|$ are shown in Table 5.3 and 5.4 respectively. In addition to the reweighting, oversampling [124] is used to account for the relative differences in the class distribution of the four jet types. Oversampling and kinematic reweighting are only applied during the training phase of a network, not during evaluation. The DNN parameters were XAVIER [125] initialised; the parameters are Gaussian distributed with $\mu = 0$ and $\sigma^2 = 1/n$ where n is the number of inputs neurons. The variance ensures that the input and output variances of one layer are identical.

During training, the network performance was tracked based on the mean accuracy. Therefore, after each epoch, the validation accuracy of the four classes was determined, and their mean value was calculated. Learning rate decay reduced the influence of the initial learning rate: if the mean accuracy on the validation did not improve for five consecutive epochs, the learning rate was reduced by a factor of two. Additionally, early stopping was implemented in order to prevent the DNN from overfitting the training sample. The training of the network was stopped after 12 consecutive epochs without an increase in the mean accuracy (i.e. the fraction of correctly classified signal and background objects), whereas the learning rate was decreased by a factor of two every five consecutive epochs without improvements. In total, the training lasted 35 epochs. The network for further analysis corresponds to the best performing training state (i.e. epoch 23).

The set of hyperparameters was optimised by comparing 800 different training configu-

rations obtained by randomly varying the hyperparameters [126] in a pre-defined range of parameter values (see Table 5.5). Depending on the hyperparameter type, the value was chosen logarithmically, linearly, or item-wise. Of the 400 best-performing network configurations, the one with the most stable signal efficiency and background rejection as a function of the jet p_T and $|\eta|$ was chosen. After selecting the network configuration, the DNN was retrained with the rounded (order of percent) set of hyperparameters. The retraining was solely aesthetically motivated and did not influence the performance. The initial hyperparameter ranges and the best performing set of hyperparameters are presented in Table 5.5.

Table 5.3.: Boundaries of the p_T -bins used as an input variable of the DNN. During the training of the network, the loss function is weighted such that p_T -bin distribution is flat.

Bin Boundaries in p_T [GeV]														
500	520	540	560	580	605	630	660	700	740	780	840	900	1000	2500

Table 5.4.: Boundaries of the $|\eta|$ -bins used as an input variable of the DNN. During the training of the network, the loss function is weighted such that $|\eta|$ -bin distribution is flat.

Bin boundaries in $ \eta $				
0.0	0.6	1.37	1.52	2.5

Table 5.5.: Summary of the network architecture and final choice of hyperparameters used in the training of the boosted $Z \rightarrow e^+e^-$ tagging algorithm. The range of parameter values used in the optimisation of the hyperparameters is shown as well.

Hyperparameter	Value	Optimisation range	Type
Number of output nodes	4	–	-
Number of hidden layers	4	[2, 6]	linear
Number of nodes per hidden layer	200	[50, 400]	linear
Number of training epochs	35	max. 100	-
Learning rate	0.002	[0.0001, 0.01]	log
Activation function	Sigmoid	[ReLU, Sigmoid, LeakyReLU]	item
Batch size	2400	[500, 5000]	linear
Dropout probability	0.01	[0, 0.6]	linear
Optimiser	Adam	–	-
Loss function	Cross e.	–	-

5.6. Performance

The performance of the DNN is evaluated by studying the confusion matrix of the four classes, the signal efficiency $\varepsilon_{Z \rightarrow ee}$ (i.e. the sensitivity), and the background rejection rates. The evaluation of the DNN performance is done on the test sample, i.e. a dataset of simulated $Z \rightarrow e^+e^-$ candidate jets that is statistically independent of the set of data used in the training process. The dataset used in the training consists of around 1.6 million $Z \rightarrow e^+e^-$ -jets, around 0.5 million $(\tau)\tau$ -jets, around 1.7 million e/γ -jets, and around 3.6 million q/g -jets.

Confusion Matrix The $Z \rightarrow e^+e^-$ identification algorithm's primary objective is to separate the signal from either of the background classes. However, the four-dimensional output offers additional freedom in constructing the final selection criterion. At first, the tagger performance is presented for each class separately, as depicted in the confusion matrix in Figure 5.22. An input jet is assigned to the class with the highest probability predicted by the DNN. The confusion matrix shows high discrimination between the signal and the three background classes and significant separation between either of the background classes. 99 % of the signal jets are correctly identified; for the background jets, this value ranges from 58 % to 85 %. The confusion occurs predominantly between the three background classes (the maximal value is 25 % for $(\tau)\tau$ identified as q/g -jets). In contrast, the confusion between the signal and each of the background classes is significantly lower (the maximal value is 1 % for $Z \rightarrow e^+e^-$ identified as e/γ -jets).

Discriminant After investigating the class-wise performance of the DNN, the probabilities of the signal and three background classes are transformed into the scalar discriminant D_{Zee} for the signal-jet hypothesis, following the arguments in Ref. [127]

$$D_{Zee} = \ln \left(\frac{f_{Z \rightarrow ee} \cdot p_{Z \rightarrow ee}}{f_{q/g} \cdot p_{q/g} + f_{e/\gamma} \cdot p_{e/\gamma} + f_{(\tau)\tau} \cdot p_{(\tau)\tau}} \right), \quad (5.7)$$

where the p_i are the probabilities for the various signal and background hypotheses as predicted by the DNN, while the f_i are the jet fractions. The optimal choice of the f_i may differ significantly between different physics analyses, depending on the expected event kinematics and background composition. For the sake of simplicity, these parameters are set in the following studies according to the jet fractions of the testing sample: i.e. $f_{Z \rightarrow ee} = 0.23$, $f_{q/g} = 0.47$, $f_{e/\gamma} = 0.23$, and $f_{(\tau)\tau} = 0.07$. Combining the probabilities of the signal and background classes allows to further optimise the algorithm's performance by tuning the relative importance of the q/g -jet, e/γ -jet and

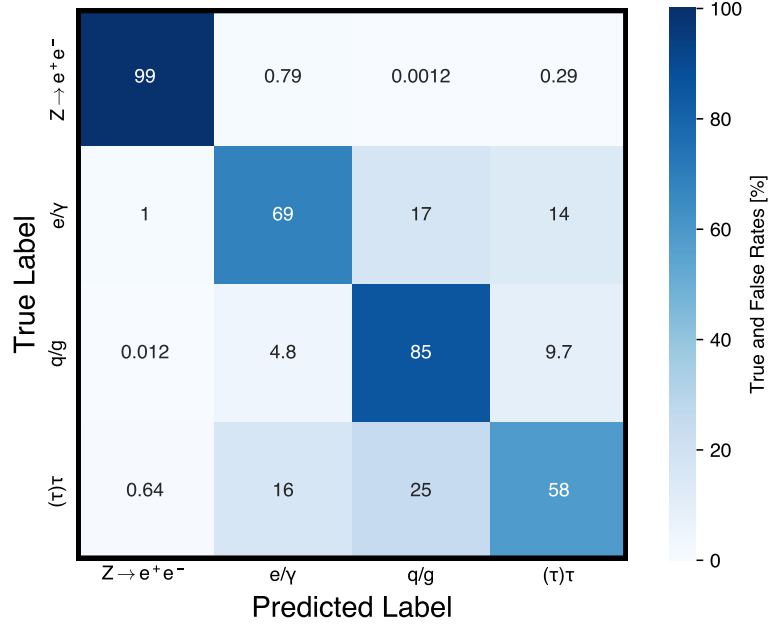


Figure 5.22.: Confusion matrix of the the signal ($Z \rightarrow e^+e^-$) or one of the three background classes (q/g , e/γ , $(\tau)\tau$). The tagger shows a high separation between signal and background and a significant separation between each of the three background classes.

$(\tau)\tau$ -jet rejection by adjusting the f_i parameters. The distribution of the final D_{Zee} score, as defined by Equation 5.7, is presented in Figure 5.23 for signal jets as well as for q/g -jets, e/γ -jets and $(\tau)\tau$ -jets. A strong separation between the signal and background sources is visible. While the D_{Zee} distribution in Figure 5.23 is based on samples from the test set, Appendix A.4 compares the D_{Zee} distribution of the train and test sample and no significant deviation for the two samples can be seen (i.e. no significant overfitting).

Depending on the selection requirement on the D_{Zee} score, the efficiencies $\varepsilon_{Z \rightarrow ee}$ and rejection rates

$$\varepsilon_{Z \rightarrow ee} = \frac{\text{True Positives}}{\text{Positives}}$$

$$\text{Rejection Rate} = \frac{\text{Negatives}}{\text{False Positives}}$$

are accessible, with the True Positives (i.e. the correctly identified signal jets), the Positives (i.e. the total number of signal jets), the Negatives (i.e. the total number of

background jets) and the False Positives (i.e. the incorrectly identified background jets). Figure 5.24 presents the receiver operating characteristics (ROC) curve. i.e the rejection rates for each background class as a function of the signal efficiency.

Operating point Operating points are defined based on selection requirements on the D_{Zee} score distribution (see Figure 5.23) ensuring a specific $Z \rightarrow e^+e^-$ identification efficiency. In order to stabilise the identification efficiency as a function of the jet p_T , the selection requirements on D_{Zee} are chosen to be p_T -dependent. Therefore, the D_{Zee} distribution is determined for different p_T intervals, as shown in Figure 5.25. For each interval, the operating point (i.e. the cut-off value on the D_{Zee} score) was determined corresponding to a defined efficiency $\varepsilon_{Z \rightarrow ee}$. An Exponential fit is applied to the operating point intervals, as depicted in Figure 5.26. A selection of various operating points and their corresponding rejection rates per background source are listed in Table 5.6.

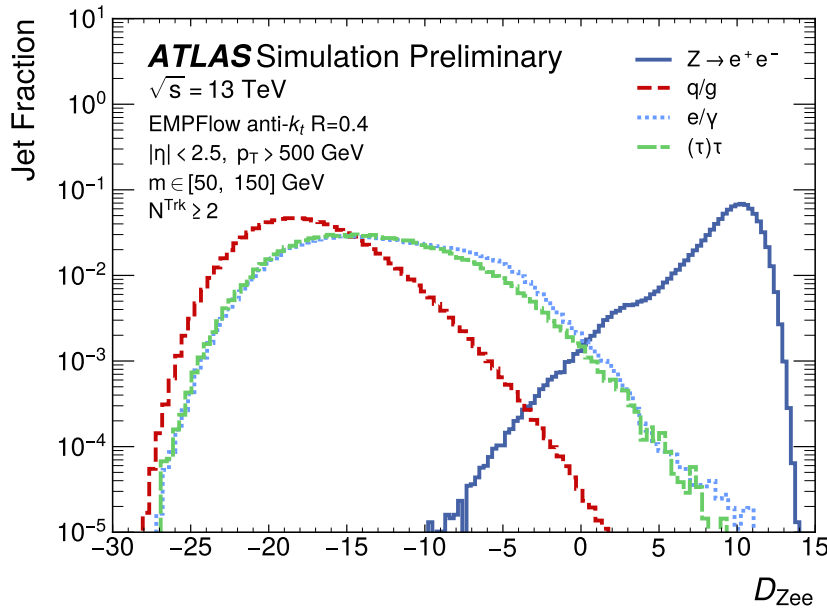


Figure 5.23.: Distribution of the discriminant of the DNN based $Z \rightarrow e^+e^-$ identification algorithm presented separately for signal jets and jets from the three background classes. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

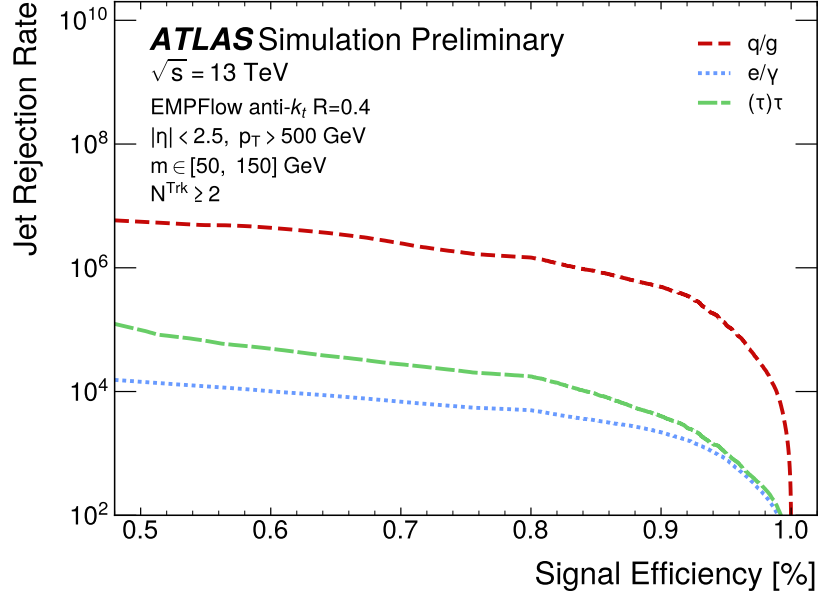


Figure 5.24.: Rejection rates for q/g -jets, e/γ -jets and $(\tau)\tau$ -jets presented as a function of the signal efficiency. Only candidates with $p_T > 500$ GeV, $|\eta| < 2.5$, and $50 \text{ GeV} < m_j < 150 \text{ GeV}$ are considered. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

Table 5.6.: Rejection rates for q/g -jets, e/γ -jets and $(\tau)\tau$ -jets for different p_T dependent operating points of the $Z \rightarrow e^+e^-$ tagging algorithm.

$\varepsilon_{Z \rightarrow ee}$	Selection	Rejection rates		
		q/g -jets	e/γ -jets	τ -jets
97%	$> 1.77 \cdot \exp\left(-0.00387\text{GeV} \cdot (p_T - 500\text{GeV})\right) + 1.21$	33591	319	379
95%	$> 1.67 \cdot \exp\left(-0.00471\text{GeV} \cdot (p_T - 500\text{GeV})\right) + 2.63$	88613	740	793
93%	$> 1.55 \cdot \exp\left(-0.00474\text{GeV} \cdot (p_T - 500\text{GeV})\right) + 3.74$	176427	1234	1443
88%	$> 1.23 \cdot \exp\left(-0.00213\text{GeV} \cdot (p_T - 500\text{GeV})\right) + 5.30$	543983	2578	4518

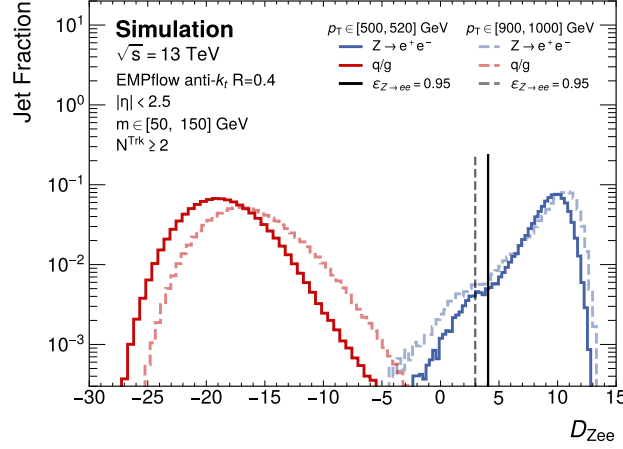


Figure 5.25.: D_{Zee} distributions of $Z \rightarrow e^+e^-$ jets (blue) and q/g jets (red) are shown for $p_T \in [500, 520]$ GeV and $p_T \in [900, 1000]$ GeV. In each p_T interval, an operating corresponding to a signal efficiency of 95 % is determined.

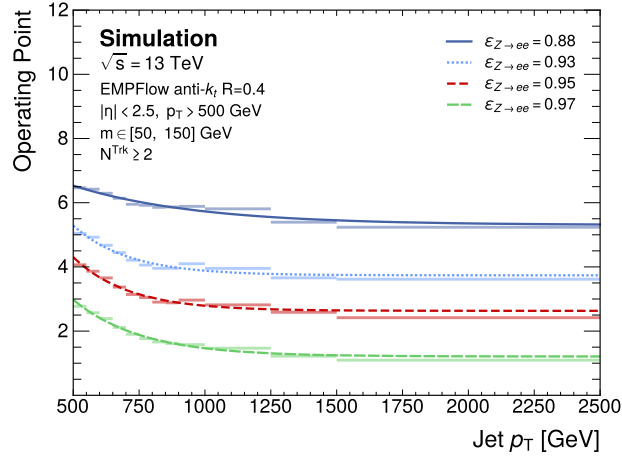


Figure 5.26.: The four operating points correspond to efficiencies $\epsilon_{Z \rightarrow ee} \in \{0.88, 0.93, 0.95, 0.97\}$. The translucent horizontal bars are the operating points corresponding to the very efficiencies in each p_T interval. An exponential fit is applied to the operating point intervals, as defined in Table 5.6. The Exponential function defines the p_T -dependent operating point (i.e. the upper/lower threshold value of the D_{Zee} distribution in Figure 5.23)

Efficiency The efficiency to reconstruct and identify boosted $Z \rightarrow e^+e^-$ decays via a single anti- k_t jet and the newly developed DNN is depicted in Figure 5.2 as a function of the angular separation between the generator-level electron-positron pair, $\Delta R_{e^+e^-}^{\text{truth}}$. The efficiency curves are presented for a selection of different operating points. Both, the reconstruction efficiency and the identification efficiencies decline with decreasing $\Delta R_{e^+e^-}^{\text{truth}}$ values. However, the new $Z \rightarrow e^+e^-$ tagging algorithm provides a significantly better reconstruction and identification performance than the standard electron reconstruction and identification techniques (see Figure 5.27). For example, for $\Delta R_{e^+e^-}^{\text{truth}}$ values between 0.06 and 0.08, the efficiency to reconstruct and identify $Z \rightarrow e^+e^-$ decays ranges from 1.2% to 4.8% (depending on the operating point) for the standard electron techniques and between 45% and 75% for the newly developed $Z \rightarrow e^+e^-$ tagging algorithm. Figure 5.28, Figure 5.29 and Figure 5.30 show the efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN is as a function of the transverse momentum p_T , the pseudorapidity η and the number of vertices N^{Vert} respectively. The signal efficiency is stable as a function of the transverse momentum and the number of vertices. In contrast, the transition region ($|\eta| \in [1.37, 1.52]$) has a degraded efficiency due to the detector's reduced performance and the high fraction of dead material.

Rejection Rate Figure 5.31 presents the background rejection rates for q/g -, e/γ -, and $(\tau)\tau$ -jets as a function the jet p_T and jet $|\eta|$ for the operating points corresponding to efficiencies of 97%, 95% and 93%, respectively. The available MC statistics are insufficient to demonstrate the p_T and $|\eta|$ dependence of the rejection rates for the 88% efficient working point. The rejection rates for all three background types vary mildly as a function of the p_T and $|\eta|$. The background rejection rates of q/g -jets (top row), e/γ -jets (middle row) and $(\tau)\tau$ -jets are around 90000, 700, and 800, respectively, for the 95% efficient operation point. The rejection rates are the largest for low p_T and $|\eta|$ values.

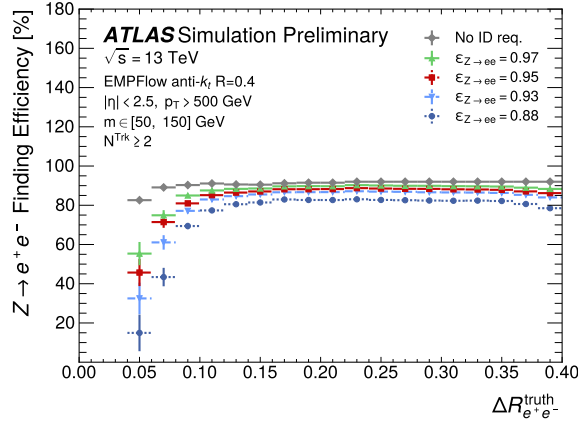


Figure 5.27.: Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the angular separation between the generator-level electron-positron pair. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0_{NNLO} PDF set.

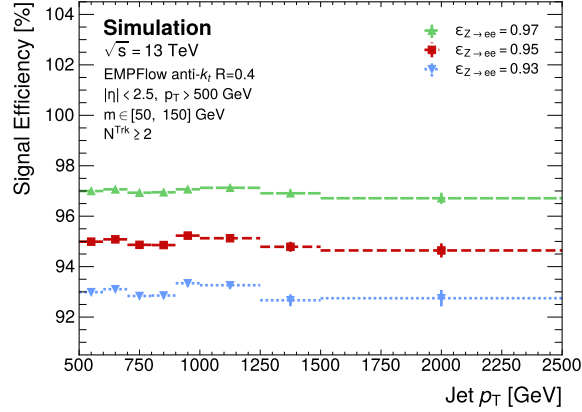


Figure 5.28.: Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the transverse momentum. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0_{NNLO} PDF set.

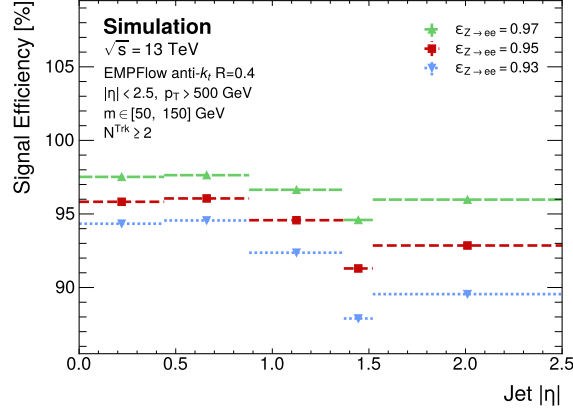


Figure 5.29.: Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the pseudorapidity. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

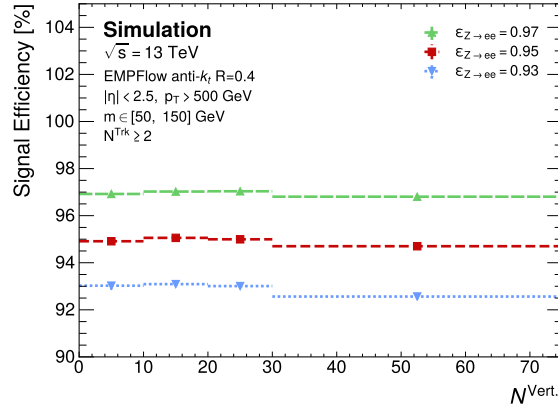


Figure 5.30.: Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the number of vertices. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

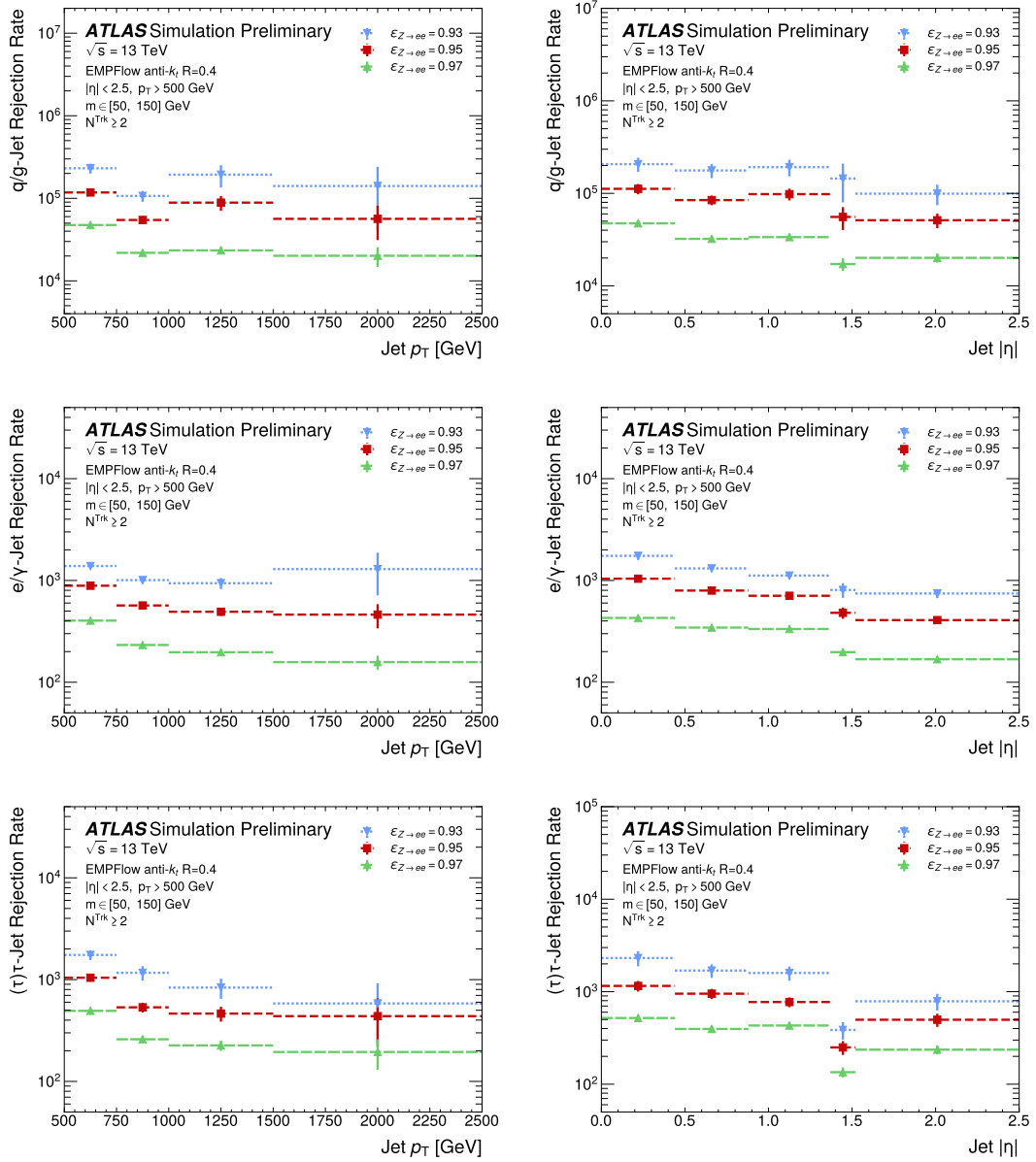


Figure 5.31.: Rejection rates for q/g -jets (top row), e/γ -jets (middle row) and $(\tau)\tau$ -jets (bottom row) presented as a function of the jet p_T (left column) and jet $|\eta|$ (right column) and for different operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

5. A new Algorithm to identify highly boosted $Z \rightarrow e^+e^-$ Decays

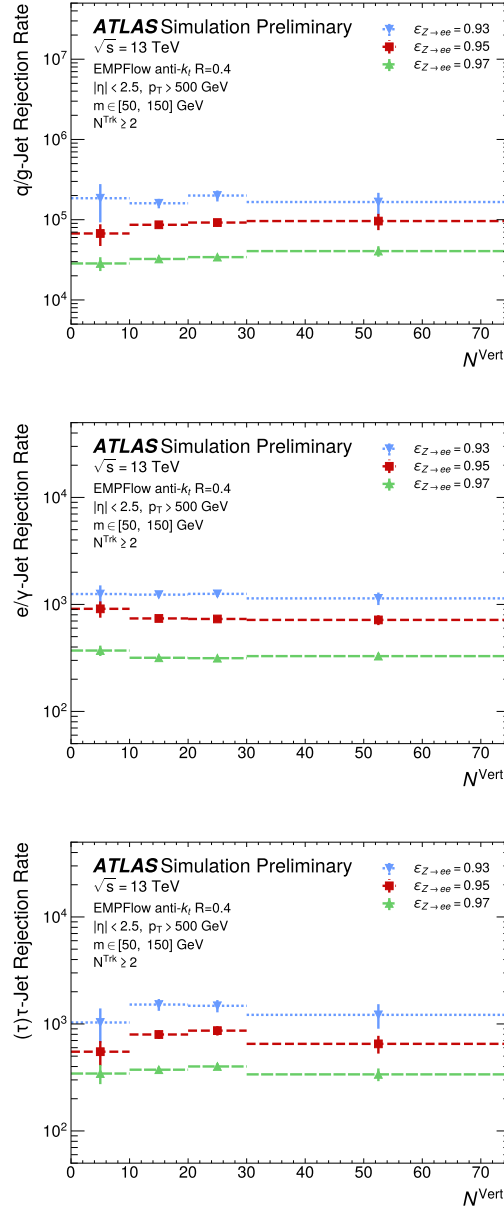


Figure 5.32.: Rejection rates for q/g -jets (top row), e/γ -jets (middle row) and $(\tau)\tau$ -jets (bottom row) presented as a function of the number interaction vertices N^{Vert} and for different operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.

5.7. Variable Importance

The importance of the DNN input observables was investigated based on the mutual information (MI) [128] of each input distribution with the DNN output. The MI is a measure of the mutual dependence between the two distributions and can therefore be used to quantify the importance of the input observables for the DNN output distribution; it is defined as

$$\begin{aligned} I(X;Y) &= D_{KL}[p(x,y)||p(x)p(y)] \\ &= \sum_{x \in X, y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \\ &= H(X) - H(X|Y), \end{aligned}$$

with the Kullback-Leibler divergence D_{KL} of the distributions $p(x,y)$ and $p(x)p(y)$, $H(X)$, the entropy of distribution X and $H(X|Y)$, the conditional entropy of X given Y . [128]. As expected, high MI values are assigned to the input variables showing a high separation between the different classes as shown in Table 5.7.

The set of input variables was optimised based on the MI as outlined in the following: at first, the network was trained with the initial set of input variables, and the mean accuracy of the best training state was determined. After removing the input variable with the lowest MI, the network was retrained, and the mean accuracy was determined for the DNN with the reduced set of inputs. This procedure was repeated until the mean accuracy degraded after removing $d\eta$ of the leading and subleading tracks and resulted in the set of 29 inputs variables as presented in Section 5.3.

The input quantities with the highest mutual information with the DNN output are the electromagnetic fraction f_{EM} , the number of tracks N^{Trk} , and the maximum fraction of energy deposited in a single layer of the EM calorimeter $\max(E_{Layer}/E_{jet})$. This result is in accordance with the input distributions as depicted in Figures 5.18, 5.19, and 5.20 since these observables show high discrimination power between the signal and each of the three background classes.

Table 5.7.: All input variables are sorted by their mutual information [128] with the DNN output in descending order. The mutual information MI quantifies the relevance of each input distribution X_i for the DNN output Y .

Input variable	MI(X_i, Y)
f_{EM}	0.606
N^{Trk}	0.582
$\max(E_{\text{Layer}}/E_{\text{Jet}})$	0.422
Planar Flow	0.396
N^{Const}	0.374
$\Delta R(t_1, t_2)$	0.338
m	0.317
Balance	0.272
$\Delta R(t_2, j)$	0.254
Leading Track $\Delta\phi$	0.227
Subleading Track $\Delta\phi$	0.156
$\Delta R(t_1, j)$	0.138
Subleading Track E / p	0.133
$r_{N=1}^{(\beta=1)}$	0.130
Width	0.111
Leading Track E / p	0.108
Q^{Sum}	0.081
Leading Track z_0	0.049
Leading Track σ_{z_0}	0.045
Leading Track d_0	0.044
Subleading Track z_0	0.040
Subleading Track σ_{z_0}	0.037
Leading Track p_e^{TRT}	0.028
Leading Track σ_{d_0}	0.026
Subleading Track p_e^{TRT}	0.025
η^{Bin}	0.025
Subleading Track σ_{d_0}	0.017
$p_{\text{T}}^{\text{Bin}}$	0.009
Subleading Track d_0	0.006

5.8. Comparison to the Standard Method

In the following, the performance of the novel $Z \rightarrow e^+e^-$ reconstructing and identification algorithm is compared to the standard method for an angular separation between the e^+e^- pair of $\Delta R_{e^+e^-}^{\text{truth}} \in [0.06, 0.08]$.

The efficiency of reconstructing the $Z \rightarrow e^+e^-$ decay via the standard method is 38.6 %, while the novel method has a reconstruction efficiency of 89.1 %. For the *Loose* operating point of the standard method, the efficiency for identifying a prompt electron with $E_T = 40$ GeV is 93 %. The efficiency of identifying $Z \rightarrow e^+e^-$ decays with an angular separation between the electrons of $\Delta R_{e^+e^-}^{\text{truth}} \in [0.06, 0.08]$, the efficiency is only 4.7 %. The novel $Z \rightarrow e^+e^-$ identification method reaches a significantly higher efficiency of 71.4 % for a global efficiency in identifying $Z \rightarrow e^+e^-$ decay of 93 % (i.e. the $Z \rightarrow e^+e^-$ decay, not one electron from the decay like for the LH based method). The $Z \rightarrow e^+e^-$ finding efficiency as function of the separation of the e^+e^- pair is depicted in Figure 5.33.

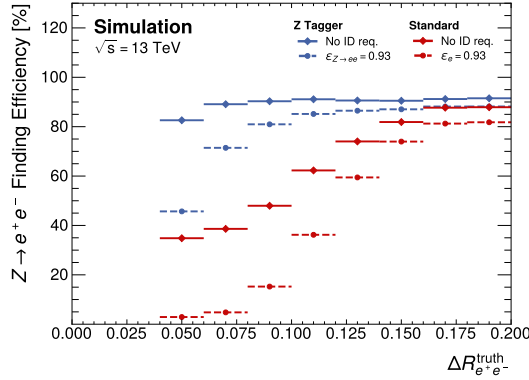


Figure 5.33.: Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the angular separation between the generator-level electron-positron pair (blue), and the efficiency to identify an electron stemming from a $Z \rightarrow e^+e^-$ decay with the standard method.

6. The Search for new heavy Resonances

This Chapter details the application of the $Z \rightarrow e^+e^-$ identification and reconstruction algorithm to a search for exotic particles such as the Z' . The analysis is carried out as a search for an excess in the invariant mass distribution, and the results are expressed as upper limits on the production cross section times branching ratio. The results obtained with the new $Z \rightarrow e^+e^-$ tagging algorithm are compared to the results of the standard method. The general search strategy is inspired by the studies presented in Reference [20].

This Chapter is structured as follows: first, the MC samples are detailed in Section 6.1. After describing the analysis strategy and event selection in Section 6.2, the statistical analysis will be explained in Section 6.3. Finally, the application of the $Z \rightarrow e^+e^-$ tagger to the search for a Z' is detailed in Section 6.4, and a comparison to the standard techniques is drawn. In this context, upper limits on the cross section times branching ratio are compared.

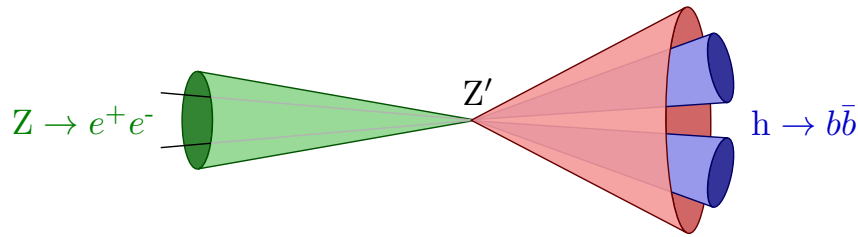


Figure 6.1.: An illustration of a hypothetical $Z' \rightarrow Zh \rightarrow e^+e^-b\bar{b}$ decay. Due to the high mass of the Z' , $m_{Z'}$, the Z boson and the Higgs boson decay are emitted back-to-back. The large transverse momenta result in a collimated topology of their decay products. The $Z \rightarrow e^+e^-$ decay is reconstructed as a small-R jet (green), and the Higgs boson decay as a large-R jet (red).

6.1. Monte Carlo Samples

The Monte Carlo (MC) simulation samples correspond to the data-taking period at the ATLAS detector between 2015 and 2018 in proton-proton (pp) collisions at $\sqrt{s} = 13$ TeV and correspond to a total integrated luminosity of 139 fb^{-1} . MC simulation samples are used to model all the major backgrounds and the signal processes. This section follows the description of the MC samples in Reference [20].

For the heavy vector triplet (HVT) interpretations, the production of Z' bosons via quark–antiquark annihilation was modelled at leading-order (LO) accuracy by MADGRAPH5 2.3.3 [129] interfaced with PYTHIA8.186 [130], which used the A14 set of tuned parameters [131] and the NNPDF2.3LO parton distribution function (PDF) set [132]. Events were generated for a range of resonance masses from 800 GeV to 5 TeV using the benchmark HVT *Model A* [33]. Higgs boson decays into $b\bar{b}$ or $c\bar{c}$ are considered, assuming the branching fractions $B(h \rightarrow b\bar{b}/c\bar{c}) = 0.569/0.0287$ and a Higgs boson mass of $m_h = 125$ GeV. The Z boson is required to decay via $Z \rightarrow \ell^+ \ell^-$, with $\ell = e$ and μ . A representative Feynman diagram is presented in Figure 6.2.

The production of W and Z bosons in association with jets was modelled by SHERPA 2.2.11 [92] with the NNPDF3.0NLO set of PDFs [133]. Diagrams with up to two additional parton emissions were simulated at NLO precision, and those with three or four additional parton emissions at LO accuracy, using the Comix [93] and OPEN Loops [94, 95, 134] libraries. Matrix elements (ME) were merged with the SHERPA parton shower [135] (PS) using the MEPS NLO prescription [98, 136–138] and the set

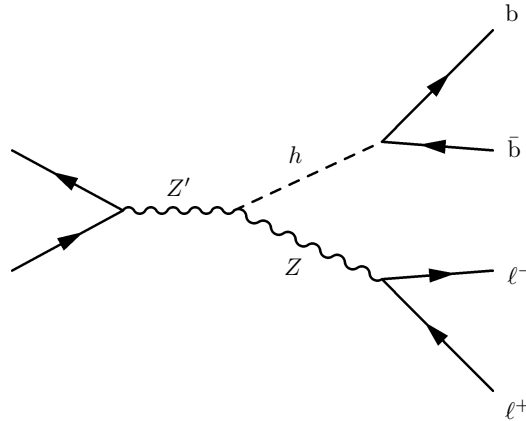


Figure 6.2.: Representative lowest order Feynman diagram for the resonant production of Z' boson via quark–antiquark annihilation.

of tuned parameters developed by the SHERPA authors. The inclusive production cross section for these samples was normalised to match a next-to-next-to-leading-order (NNLO) prediction [139].

The production of top-quark pair ($t\bar{t}$) events was modelled using the POWHEG Box v2 [140–143] generator. The matrix elements were calculated at NLO precision in QCD using the five-flavour scheme and the NNPDF3.0_{NLO} PDF set, assuming a top-quark mass of 172.5 GeV. The PS, hadronisation, and underlying event were modelled with the PYTHIA 8.230 [144] generator using the A14 set of tuned parameters and the NNPDF2.3_{LO} set of PDFs. The h_{damp} parameter¹ was set to $1.5 m_{top}$ [145]. The cross section was calculated at NNLO precision, including the resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms with TOP++ 2.0 [146–152].

Diboson events (WW , WZ , ZZ) with semileptonic decays were simulated using SHERPA 2.2.1 with the NNPDF3.0_{NNLO} PDF set, including off-shell effects and Higgs boson contributions where appropriate. Diagrams with up to one additional emission were calculated at NLO accuracy in QCD, while diagrams with two or three parton emissions were described at LO accuracy [153]. They were merged and matched using the MEPS NLO prescription. Loop-induced diboson processes that are initiated via the gg production mode were simulated at LO in QCD for diagrams with up to one additional parton emission in the matrix element using OPENLOOPS in SHERPA 2.2.1, which used the NNPDF3.0_{NNLO} PDF set. Diboson events with fully leptonic decays were modelled using SHERPA2.2.1 and the NNPDF3.0_{NNLO} PDF set.

Finally, the production of an SM Higgs boson in association with a vector boson was simulated using POWHEG Box v2, interfaced with PYTHIA8.212 for PS and non-perturbative effects. The POWHEG prediction is accurate to next-to-leading order for production of Vh plus one jet. The loop-induced $gg \rightarrow Zh$ process was generated separately at LO. The PDF4LHC15 PDF set [154] and the AZNLO set of tuned parameters [155] of PYTHIA8.212 were used. The $gg \rightarrow Zh$ production cross section was calculated at NLO precision, including the resummation of next-to-leading logarithmic (NLL) soft gluon terms [156]. For the generation of Vh events, the Higgs boson mass was set to 125 GeV.

All simulated event samples include the effect of multiple pp interactions in the same and neighbouring bunch crossings (pile-up) by overlaying simulated minimum-bias events on each generated signal or background event.

The minimum-bias events were simulated with the single-, double- and non-diffractive pp processes of PYTHIA 8.186 using the A3 set of tuned parameters [104] and the

¹The h_{damp} parameter is a resummation damping factor and one of the parameters that control the matching of POWHEG matrix elements to the PS, effectively regulating the high- p_T radiation against which the $t\bar{t}$ system recoils.

NNPDF2.3LO PDF. For all MADGRAPH and POWHEG Box samples, the EVTGEN 1.6.0 program was used for the bottom and charm hadron decays. The generated samples were processed using the GEANT4-based ATLAS detector simulation [102, 103] and the simulated events were reconstructed using the same algorithms as were used for the data events.

Simulated events were corrected to compensate for differences between data and simulations regarding the energy (or momentum) scale and resolution of leptons and jets, the efficiencies for the reconstruction, identification, isolation and triggering of leptons, and the tagging efficiency for heavy-flavour jets.

6.2. Analysis Strategy and Event Selection

The $Z' \rightarrow Zh \rightarrow \ell^+ \ell^- b \bar{b}$ signal region (SR) is defined as described in Table 6.1. $Z \rightarrow e^+ e^-$ decays are identified in two different ways: by the standard method of reconstructing and identifying the $Z \rightarrow e^+ e^-$ decay by resolving the two electrons, and by using the novel $Z \rightarrow e^+ e^-$ reconstruction and identification as detailed in Section 5.3.

Electrons are reconstructed from ID tracks matched to energy clusters in the electromagnetic calorimeter and originating from the primary vertex (PV). The latter condition is satisfied by a requirement on the transverse impact parameter significance, $|d_0|/\sigma(d_0) < 5.0$, and on the longitudinal impact parameter, $|z_0 \sin(\theta)| < 0.5$ mm. Electrons must satisfy requirements for the electromagnetic shower shape, track quality, and track-cluster matching using a likelihood-based approach. The *Loose* operating point is used [157]. Electrons have to be isolated from hadronic activity in the event: both the transverse energy sum and the scalar sum of transverse momenta of all ID tracks within a cone of variable size around the electron have to be smaller than 0.06 times the electron's transverse energy E_T . The maximum cone size is $\Delta R = 0.2$, shrinking for larger E_T [157]. Electron candidates must have a minimum p_T of 27 GeV and lie within $|\eta| < 2.47$.

Muons [69] are identified by matching tracks found in the ID to either energy deposits in the calorimeter system consistent with a minimum-ionising particle (*calorimeter-tagged muons*) or full tracks (*combined muons*) or track segments (*segment-tagged muons*) reconstructed in the muon spectrometer. Muons reconstructed as standalone tracks in the muon spectrometer are also considered. Like electrons, they have to fulfil impact-parameter requirements: $|d_0|/\sigma(d_0) < 3.0$ and $|z_0 \sin(\theta)| < 0.5$ mm. Muons are required to pass the *Medium* identification operation point. Muons must also be isolated: in the ID system, the p_T sum within a variable-size cone around the combined track

has to be smaller than 0.06 times the muon's transverse momentum. The maximum cone size is $\Delta R = 0.3$, shrinking for larger muon p_T . Muon candidates are required to have a minimum p_T of 27 GeV and to lie within $|\eta| < 2.5$.

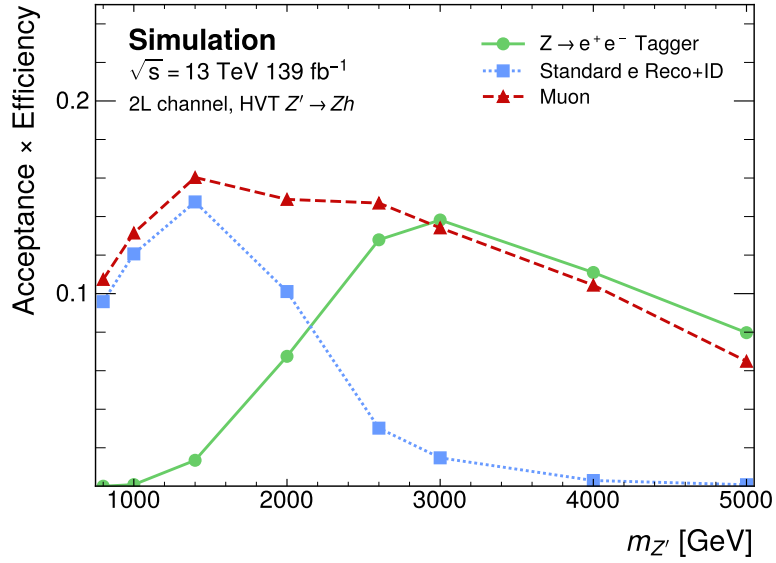
Three different signal regions (SR) are defined: the standard electron SR, the muon SR and the $Z \rightarrow e^+e^-$ SR. In the standard electron and muon SR, the Z boson is reconstructed from the sum of the four-momenta of the two charged leptons. In the $Z \rightarrow e^+e^-$ region, the $Z \rightarrow e^+e^-$ decay is reconstructed as a jet, clustered via the anti- k_t algorithm [74] using a radius parameter of 0.4 and particle-flow objects as inputs. The reconstructed Z boson and Higgs boson must have a transverse momentum of $p_T > 250$ GeV in the standard electron and muon region. In the $Z \rightarrow e^+e^-$ region the selection on the transverse momentum is defined as $p_T > 500$ GeV. The mass-window of Z boson candidates is $m^Z \in [50, 150]$ GeV in all SR.

The Higgs boson is reconstructed as a large-R jet. The identification is based on the double b -tagging algorithm: it identifies boosted $H \rightarrow b\bar{b}$ decays by combining the kinematics of one large-R jet with flavour tagging information and properties of up to three associated track jets with a DNN. A detailed description can be found in Reference [28]. The mass-window of Higgs boson candidates is $m^h \in [80, 150]$ GeV.

Finally, only events fulfilling $\min(p_T^Z, p_T^h) / m^{Zh} > 0.35$ are selected. The products of acceptance and reconstruction efficiency corresponding to this selection are presented in Figure 6.3 as a function of the resonance mass $m_{Z'}$. For a resonance mass $m_{Z'}$ of 3000 GeV, the $Z \rightarrow e^+e^-$ tagger reaches ten times higher values than the standard method; for resonance masses $m_{Z'}$ of 5000 GeV, the ratio reaches a value of 100. Event distributions of m_{Zh} for the $Z \rightarrow e^+e^-$ SR are presented in Figure 6.4.

Table 6.1.: A list of the selection criteria in three signal regions included in the statistical analysis of the Z' model hypotheses.

	$Z \rightarrow e^+e^-$ Tagger	Electron Reco+ID	Muon Reco+ID
Z candidate	1 small-R jet Z tagged	2 electrons $\eta < 2.47$ $p_{\text{T}} > 27$ GeV <i>Loose</i> OP	2 muons $\eta < 2.5$ $p_{\text{T}} > 27$ GeV <i>Medium</i> OP
Higgs candidate	Higgs-tagged large-R jet [28]		
p_{T}^Z	> 500 GeV	> 250 GeV	
p_{T}^h	> 500 GeV	> 250 GeV	
m^Z	$\in [50, 150]$ GeV		
m^h	$\in [80, 150]$ GeV		
$\min(p_{\text{T}}^Z, p_{\text{T}}^h) / m^{Zh}$	> 0.35		


 Figure 6.3.: Product of acceptance and selection efficiency for $Z' \rightarrow Zh \rightarrow \ell^+\ell^-b\bar{b}$ as a function of the resonance mass $m_{Z'}$.

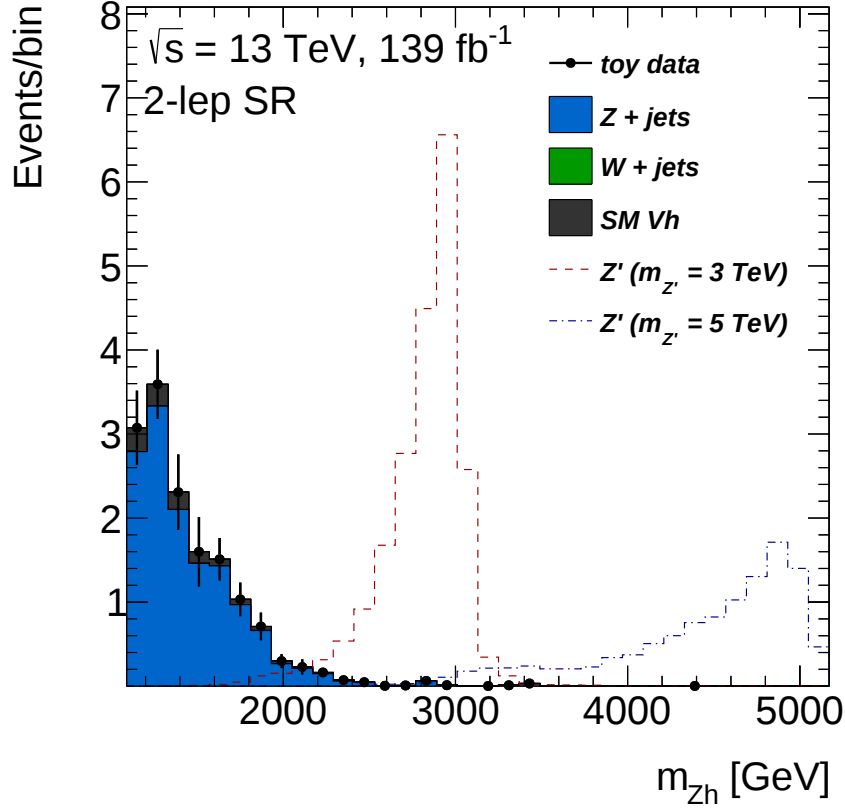


Figure 6.4.: Event distributions of m_{Zh} for the $Z \rightarrow e^+e^-$ SR. The vertical axis shows the number of data events divided by the bin width. The signal for the benchmark HVT Model A with $m_{Z'} = 3 \text{ TeV}$ (red) and $m_{Z'} = 5 \text{ TeV}$ (blue), are normalised to 1 fb^{-1} , and are shown as dashed lines.

6.3. Statistical Analysis

Expected upper limits at the 95 % confidence level (CL) on the product of the cross section for $pp \rightarrow Z'$ and the branching fraction to Zh for the $\ell^+\ell^-b\bar{b}$ final state are determined. The expected upper limits are accessible by a simultaneous profile likelihood fit on the expected signal and background distributions of the reconstructed Z' mass [158].

The number of events in bin b of specific channel c is given by

$$s_{cb} = s_{\text{tot}} \int_{cb} f_s(x; \theta_s) dx, \text{ and}$$

$$b_{cb} = b_{\text{tot}} \int_{cb} f_b(x; \theta_b) dx,$$

where s_{tot} and b_{tot} are the total number of signal events and background events respectively. The f_s and f_b are the probability density functions representing the shape of the $m_{Z'}$ distribution, that are characterised by the parameters $\theta = \{\theta_s, \theta_b\}$. The probability distribution of one bin b in a specific channel c can be expressed by the Poisson distribution

$$\text{Pois}(n_{cb}; \mu s_{cb} + b_{cb}) = \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)},$$

which depends on the number of observed events n_{cb} for the expectation value given by $\mathbb{E}[n_{cb}] = \mu s_{cb} + b_{cb}$. The signal strength μ connects the MC signal cross section and the signal cross section in data (i.e. $\sigma_s^{\text{data}} = \mu \cdot \sigma_s^{\text{MC}}$). Then, by determining the product of the Poisson distributions in each bin and channel, the Likelihood (LH)

$$L(\mu, \theta) = \prod_c \prod_b \text{Pois}(n_{cb}; \mu s_{cb} + b_{cb}) \prod_m \zeta_m(\theta) \quad (6.1)$$

is accessible. The ζ_m are constraints on the nuisance parameters θ determined in auxiliary measurements. For testing a hypothesised value of the signal strength μ , the profile LH ratio

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

is determined. $\hat{\hat{\theta}}$ are the maximum-likelihood (ML) estimators for a given μ . $\hat{\mu}$ and $\hat{\theta}$ are the ML estimators of the unconditional L . The profile LH ratio will be close to one, if the measurement and the expectation from the hypothesis are in agreement.

The test statistic

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

is constructed. A value of zero indicates maximal agreement between measurement and expectation while the value increases with decreasing agreement. Via the probability density function of the test statistic f , the p-value

$$p_\mu = \int_{q_\mu}^{\infty} f(q_{\mu'} | \mu) dq_{\mu'},$$

is accessible. Finally, the value of μ is determined, which corresponds to $p_\mu = 0.05$, i.e. a confidence level (CL) of 0.95. That value is the upper limit on the signal strength μ_{up} . [158–160]

6.4. Upper Limits on the Cross Section times Branching Ratio

The expected upper limits on the production cross section times branching ratio $\sigma(Z') \times \text{BR}(Z' \rightarrow Zh)$ as a function of the resonance mass $m_{Z'}$ are shown in Figure 6.5. The exclusion limits are shown separately for Z bosons reconstructed and identified with the standard method (red) and the novel $Z \rightarrow e^+e^-$ algorithm (black). Both analyses include events with two muons. The green and yellow bands indicate the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty of the exclusion limits determined with the novel $Z \rightarrow e^+e^-$ algorithm. For resonance masses $m_{Z'}$ of 2 TeV, 3 TeV and 5 TeV, the exclusion limits determined with the standard method of reconstructing electrons are 2.6 fb, 2.0 fb, and 3.7 fb. With values of 1.9 fb, 1.1 fb and 1.8 fb, the exclusion limits determined with the $Z \rightarrow e^+e^-$ tagger are significantly tighter. The tightest exclusion limit is reached for resonance masses $m_{Z'}$ of 3000 GeV as the approach using the $Z \rightarrow e^+e^-$ tagger has a maximum of the acceptance $\times$ efficiency (see Figure 6.3).

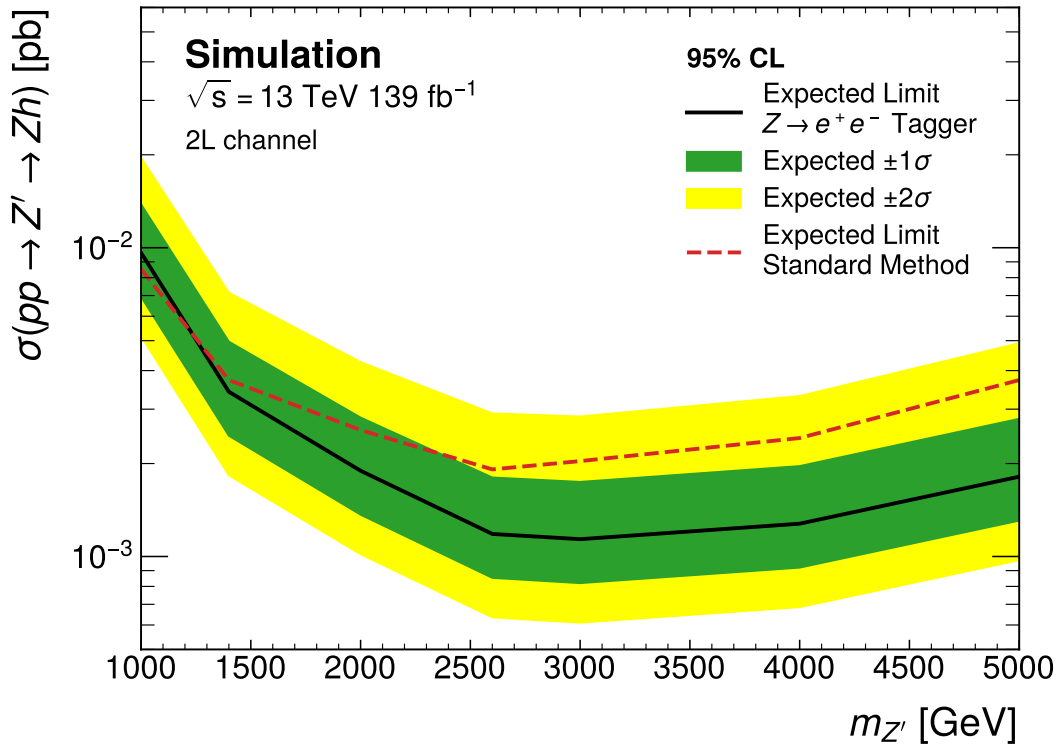


Figure 6.5.: Upper limits at the 95 % CL on the product of the cross section for $pp \rightarrow Z'$ and the branching fraction to Zh from the 2-lepton (2L) channels, separately for the reconstruction and identification via the standard method (red), and the $Z \rightarrow e^+e^-$ tagger (black). Both signal regions include muons.

7. Concluding Remarks

Many beyond the standard model theories predict the existence of new heavy particles that predominantly decay into pairs of bosons. This thesis details the development and evaluation of a new algorithm dedicated to identifying highly boosted $Z \rightarrow e^+e^-$ decays.

Jets, clustered via the anti- k_t algorithm using a radius parameter of 0.4 and particle-flow objects as inputs, are used to reconstruct the $Z \rightarrow e^+e^-$ candidates. The response scale and resolution of the mass and transverse momentum are studied for $Z \rightarrow e^+e^-$ jets in different jet p_T and jet $|\eta|$ intervals. The jet mass and momentum response scale are around unity for most of the probed kinematic range, while the mass and momentum resolutions range from 2% to 5% and 1% to 2%, respectively.

A deep neural network trained on the properties of the jet, including cluster and track information, is used to reduce the backgrounds. The $Z \rightarrow e^+e^-$ identification algorithm is trained on samples of the signal class (i.e. $Z \rightarrow e^+e^-$) and three background classes (i.e. q/g , e/γ , $(\tau)\tau$). A multidimensional output offers high flexibility when constructing the network discriminant. Signal efficiencies and background rejection rates are evaluated as a function of the jet properties. The $Z \rightarrow e^+e^-$ identification algorithm shows high performance for a large kinematic range. The background rejection rates of q/g -, e/γ -, and $(\tau)\tau$ -jets are around 90000, 700, and 800, respectively, for the 95% efficient operation point.

The newly developed $Z \rightarrow e^+e^-$ tagging algorithm provides significantly higher reconstruction and identification efficiencies than the standard ATLAS techniques for electron reconstruction and identification when applied to highly collimated e^+e^- systems originating from boosted Z boson decays. Therefore, this $Z \rightarrow e^+e^-$ tagging algorithm promises to be a useful tool to enhance the sensitivity to searches for diboson resonances in final states, including leptonic decays of highly boosted Z bosons such as $Z' \rightarrow ZZ \rightarrow \ell\ell qq$ [19], $Z' \rightarrow Zh \rightarrow \ell\ell bb$ [20] or $W' \rightarrow WZ \rightarrow \ell\nu\ell\ell$ [21].

Finally, the novel approach was compared to the standard method in terms of expected upper limits on the production cross section times branching ratio in the search for a hypothetical new heavy vector boson Z' . The exclusion limits of the standard method are 2.6 fb, 2.0 fb, and 3.7 fb for resonance masses $m_{Z'}$ of 2 TeV, 3 TeV and 5

TeV, respectively. With values of 1.9 fb, 1.1 fb and 1.8 fb, the novel method reaches significantly tighter exclusion limits for the same resonance masses.

The jet energy reconstruction is expected to improve significantly by deriving pile-up and MC-based corrections as included in JES corrections dedicated to the $Z \rightarrow e^+e^-$ jets. Furthermore, applying the GSF-refitting procedure of the standard electron tracks would improve the track quality significantly. However, the significantly mitigated reconstruction efficiency of the GSF tracks for low angular separation between the e^+e^- pairs needs to be understood.

Many applications in particle physics research rely on outdated object identification, such as Likelihood-based methods, to this day. However, modern machine learning-based approaches offer to improve particle research in a wide range of fields. One example is the presented study since the $Z \rightarrow e^+e^-$ identification algorithm based on a DNN performs significantly better than the Likelihood-based approach. Stepping further in the direction of machine learning might offer further improvement. Powerful b-tagging algorithms such as MV2 [161] or DL1 [161] combine information from low-level taggers, exploiting reconstructed track and vertex information, into machine learning classifiers. A similar approach might be suitable for the $Z \rightarrow e^+e^-$ identification algorithm: the track information could be processed by a recurrent neural network (RNN), while the jet constituents are processed by a convolutional neural network (CNN) [162]. Finally, the outputs of these two low-level taggers are combined via a DNN.

The generic approach of reconstructing merged di-lepton final states as jets and reducing the backgrounds with a DNN can easily be adapted for a more general approach. For example, next to searches for new heavy resonances, highly collimated e^+e^- pairs are expected in searches for exotic light particles such as *dark photons* [163]. This could be achieved by a mass-decorrelated tagger [164] developed with an adversarial training approach. Next to the feature extractor and the classifier separating $Z \rightarrow e^+e^-$ jets from the relevant backgrounds, a mass prediction network could be added. A high accuracy of the prediction of the jet mass from the learned features corresponds to a strong correlation of the features with the jet mass. Therefore, the mass prediction would be included in the loss function as a penalty for minimising the joint loss. Furthermore, the approach is not limited to electrons; a similar strategy could be developed for the reconstruction and identification of $X \rightarrow \mu^+\mu^-$ decays.

A. Appendix

A.1. Track Reconstruction Studies

Inspired by the standard electron identification and reconstruction, inner detector tracks refitted with Gaussian sum filters, as described in Section 3.3.1, were tested. However, the GSF tracks show a significantly limited reconstruction efficiency, as depicted in Figure A.1.

Initial studies were performed to understand the degraded performance of the GSF tracks. However, an in-detail investigation of the underlying algorithms will be needed to solve this shortcoming. Therefore, the $Z \rightarrow e^+e^-$ identification algorithm relies on the standard inner detector tracks, i.e. tracks not refitted with Gaussian sum filters.

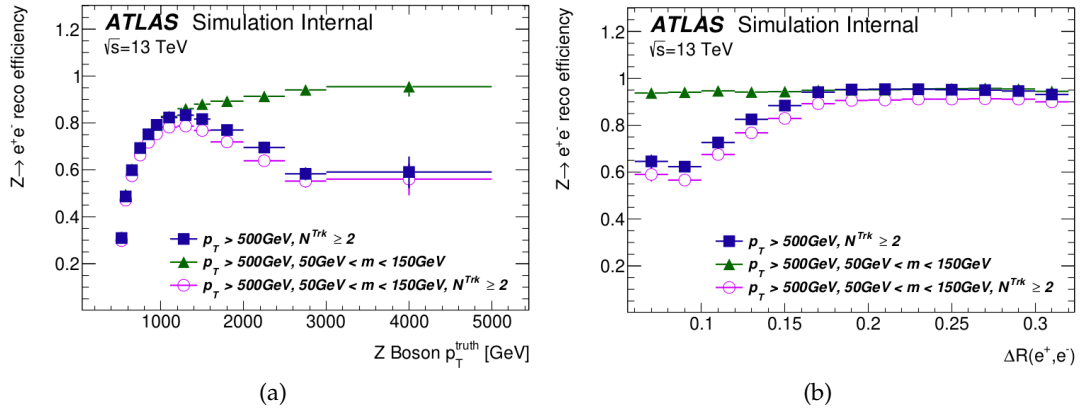


Figure A.1.: The $Z \rightarrow e^+e^-$ reconstruction is significantly reduced if at least two GSF tracks are required.

A.2. Jet Resolution Studies

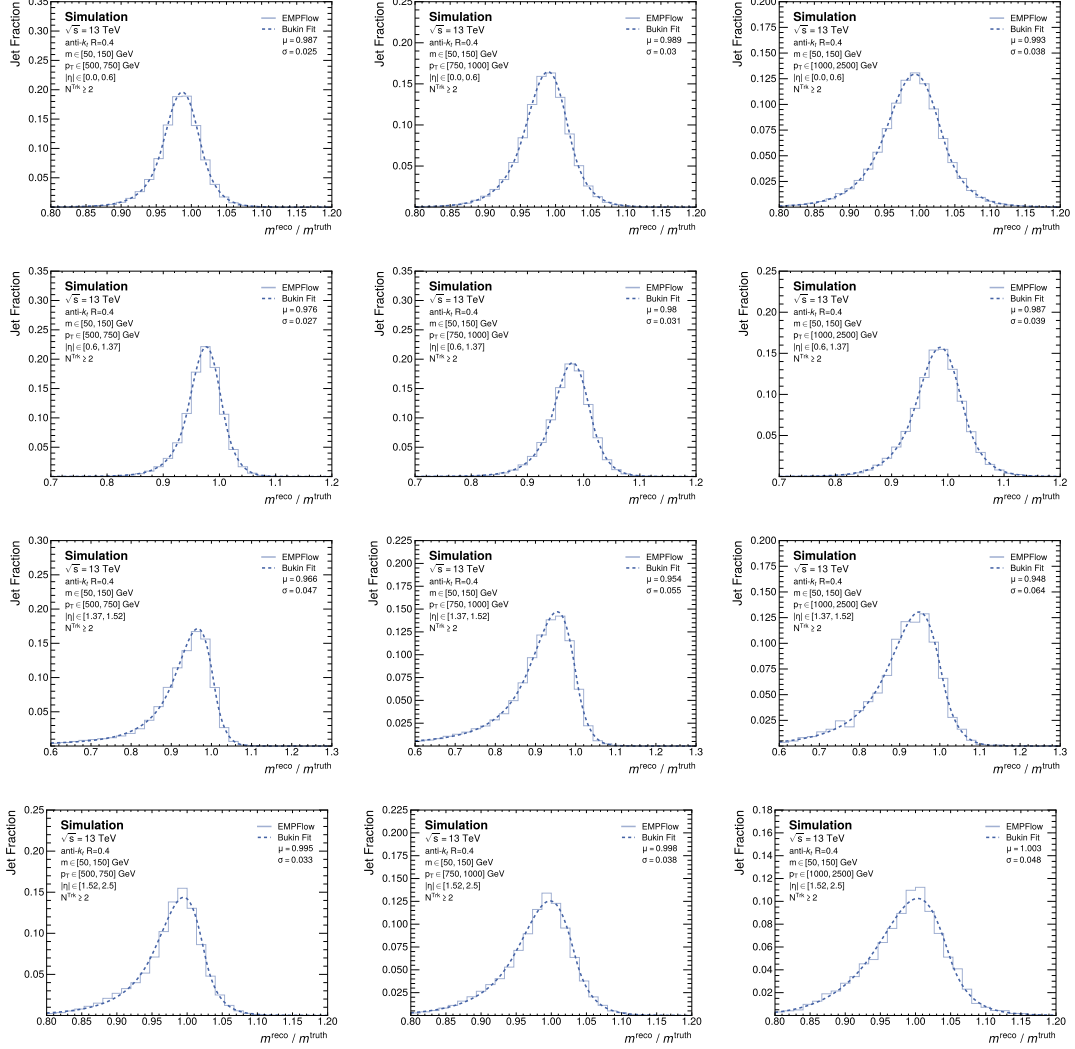


Figure A.2.: Mass response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPflow without JES for different event p_T and $|\eta|$ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.

A. Appendix

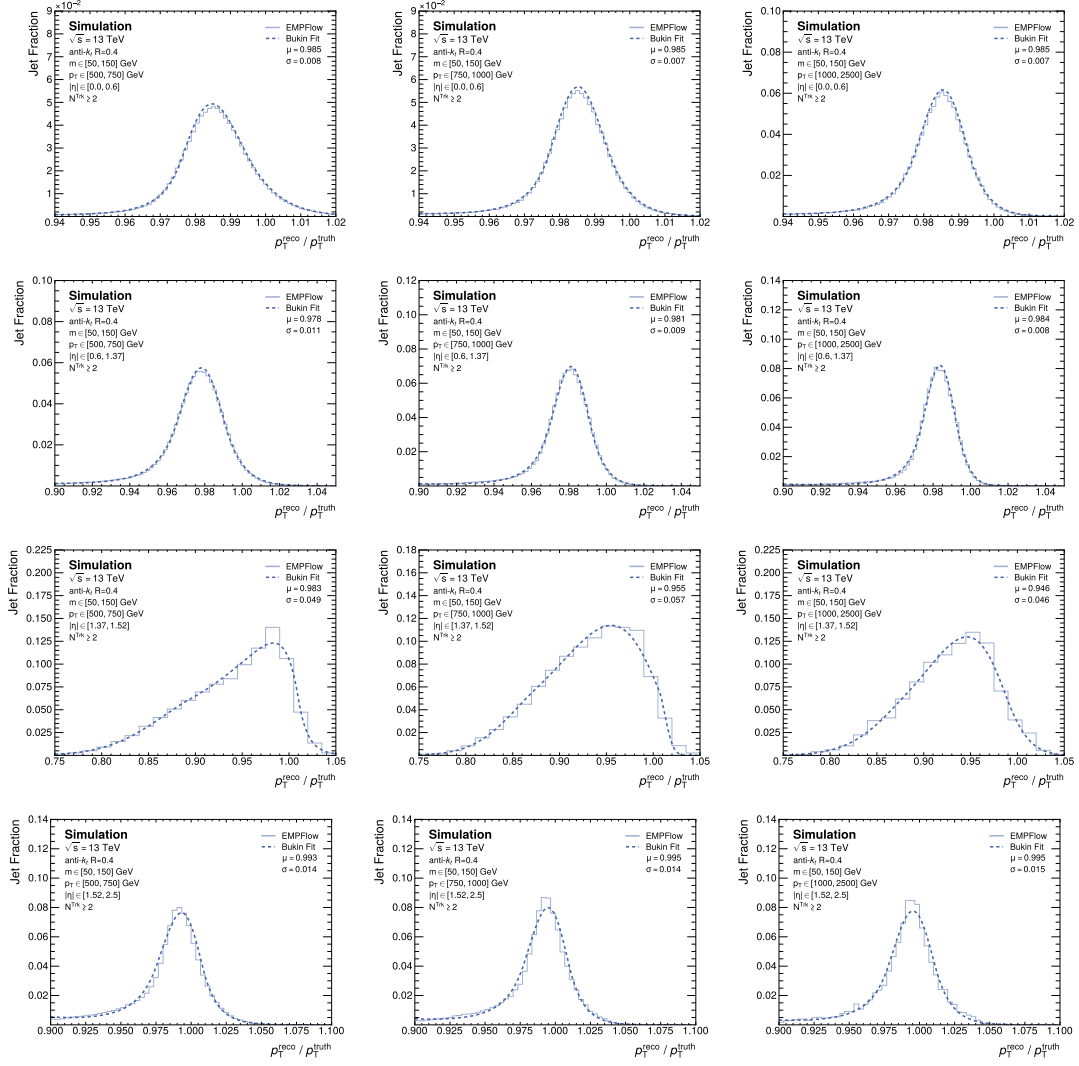


Figure A.3.: Transverse momentum response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPflow without JES for different p_T and $|\eta|$ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.

A. Appendix

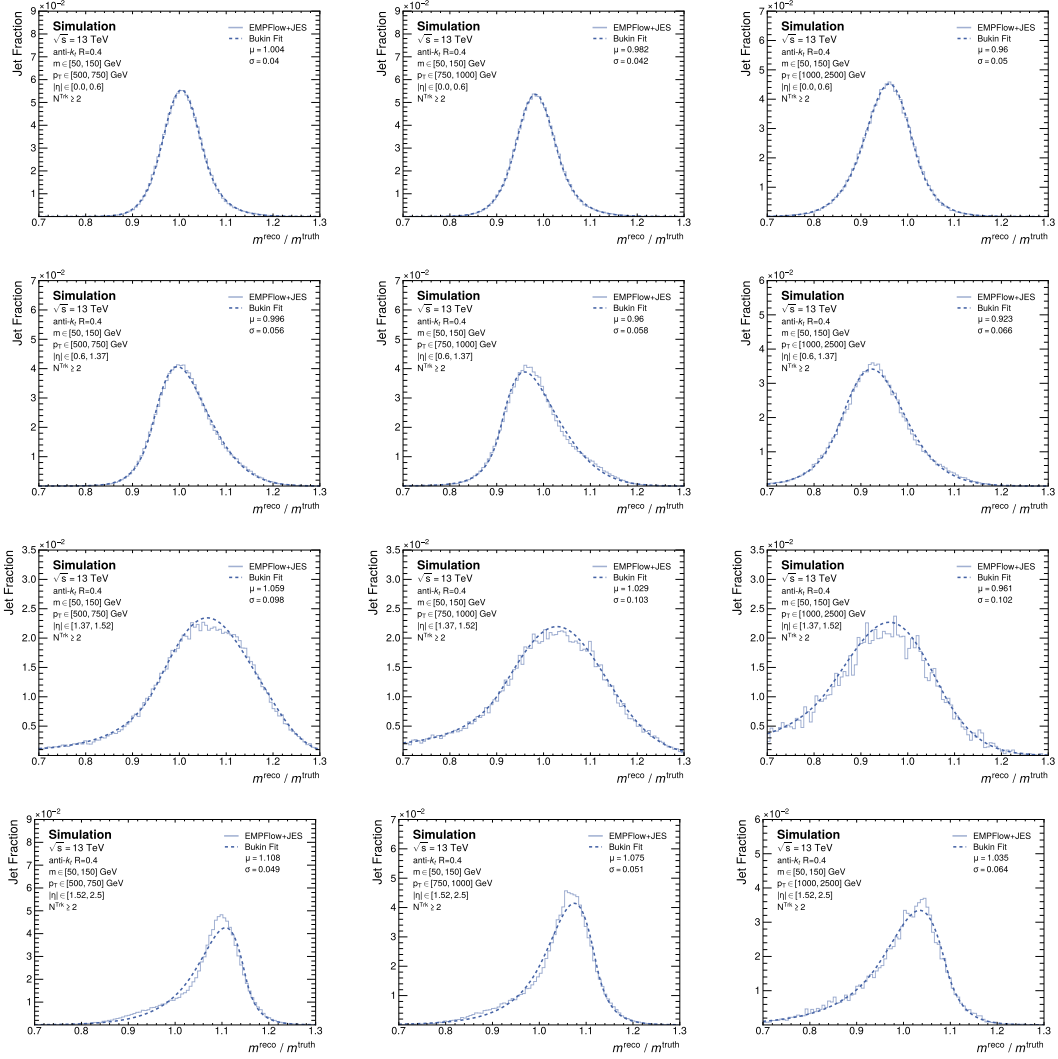


Figure A.4.: Mass response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPFlow+JES for different p_T and $|\eta|$ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.

A. Appendix

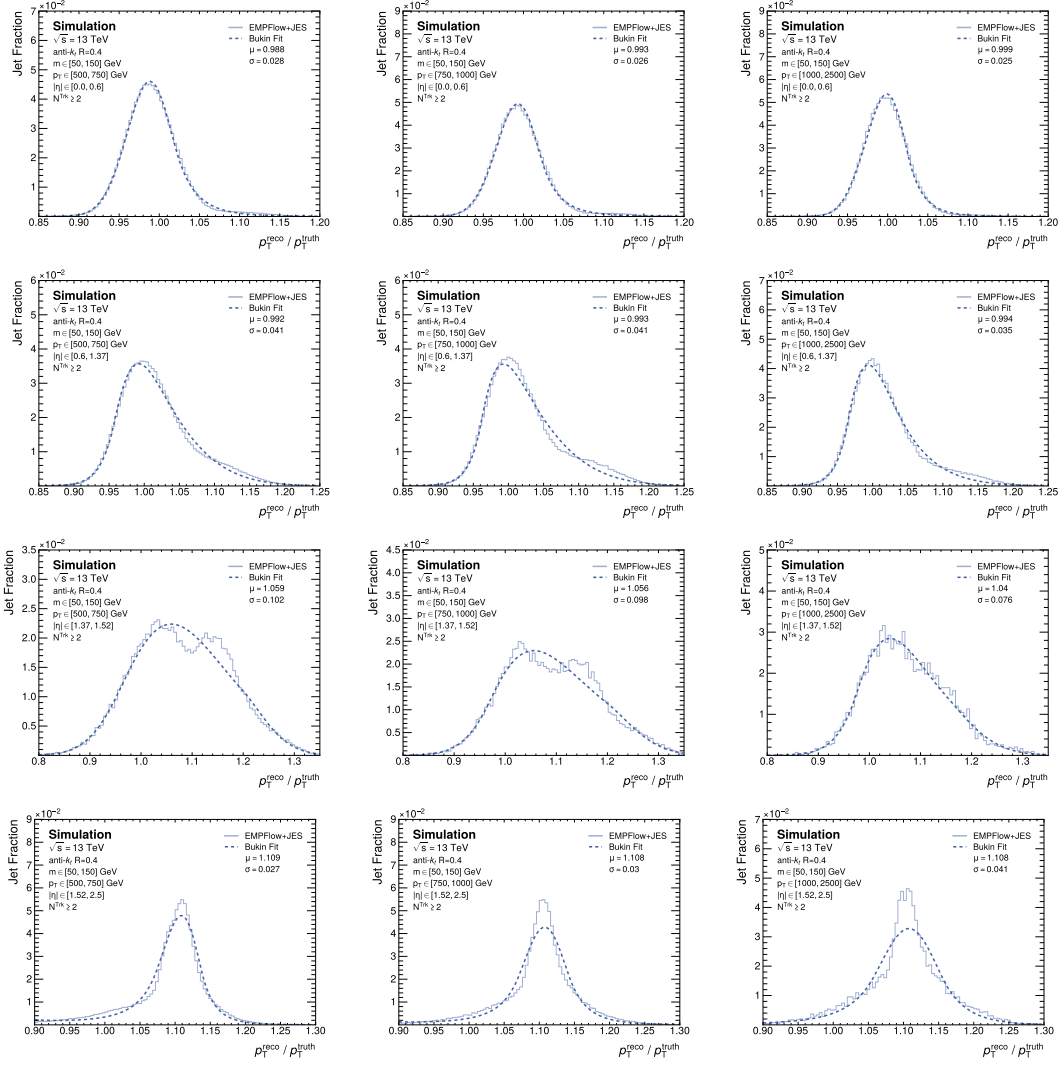


Figure A.5.: Transverse momentum response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPflow+JES for different p_T and $|\eta|$ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.

A. Appendix

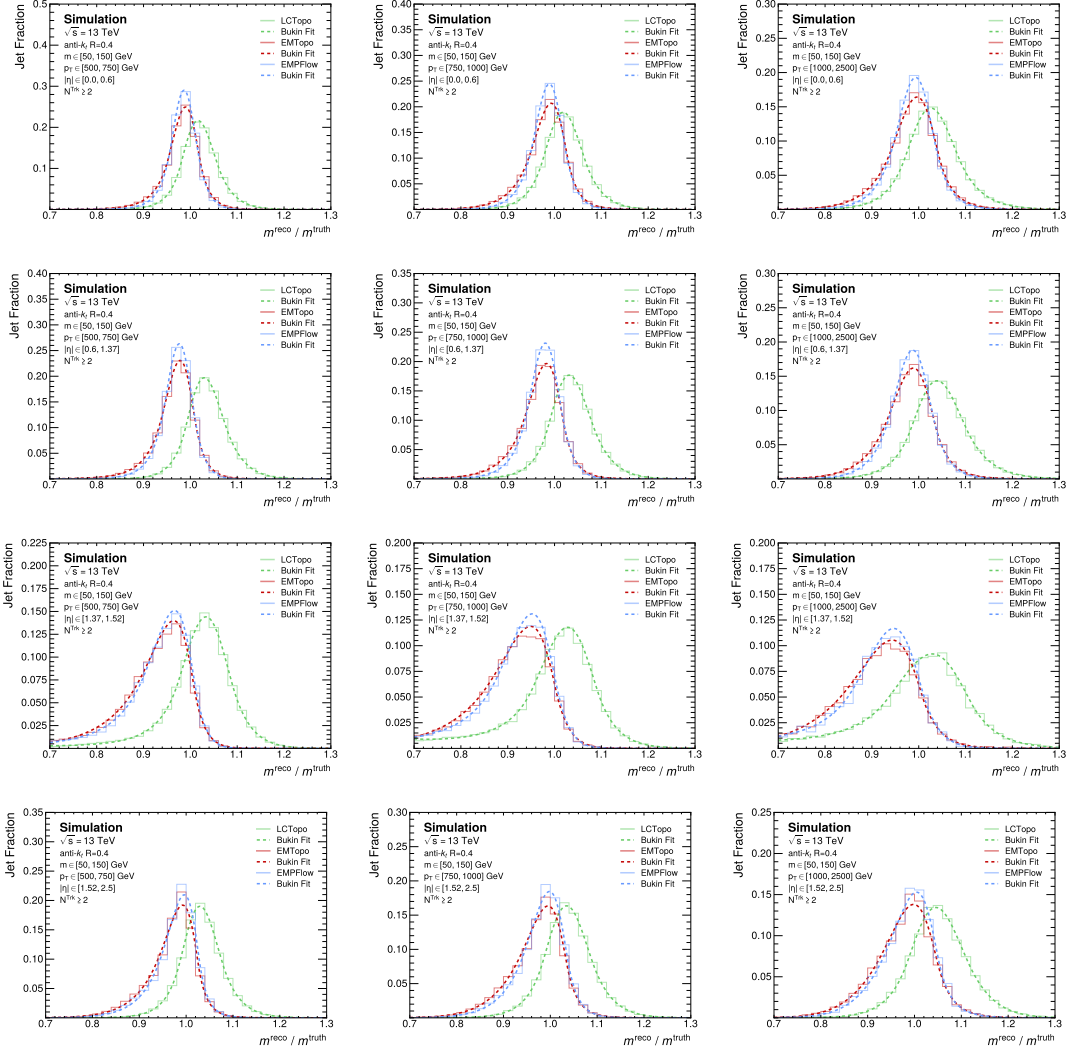


Figure A.6.: Mass response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with LCTopo (green), EMTopo (red) and EMPFlow (blue) for different p_T and $|\eta|$ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.

A. Appendix

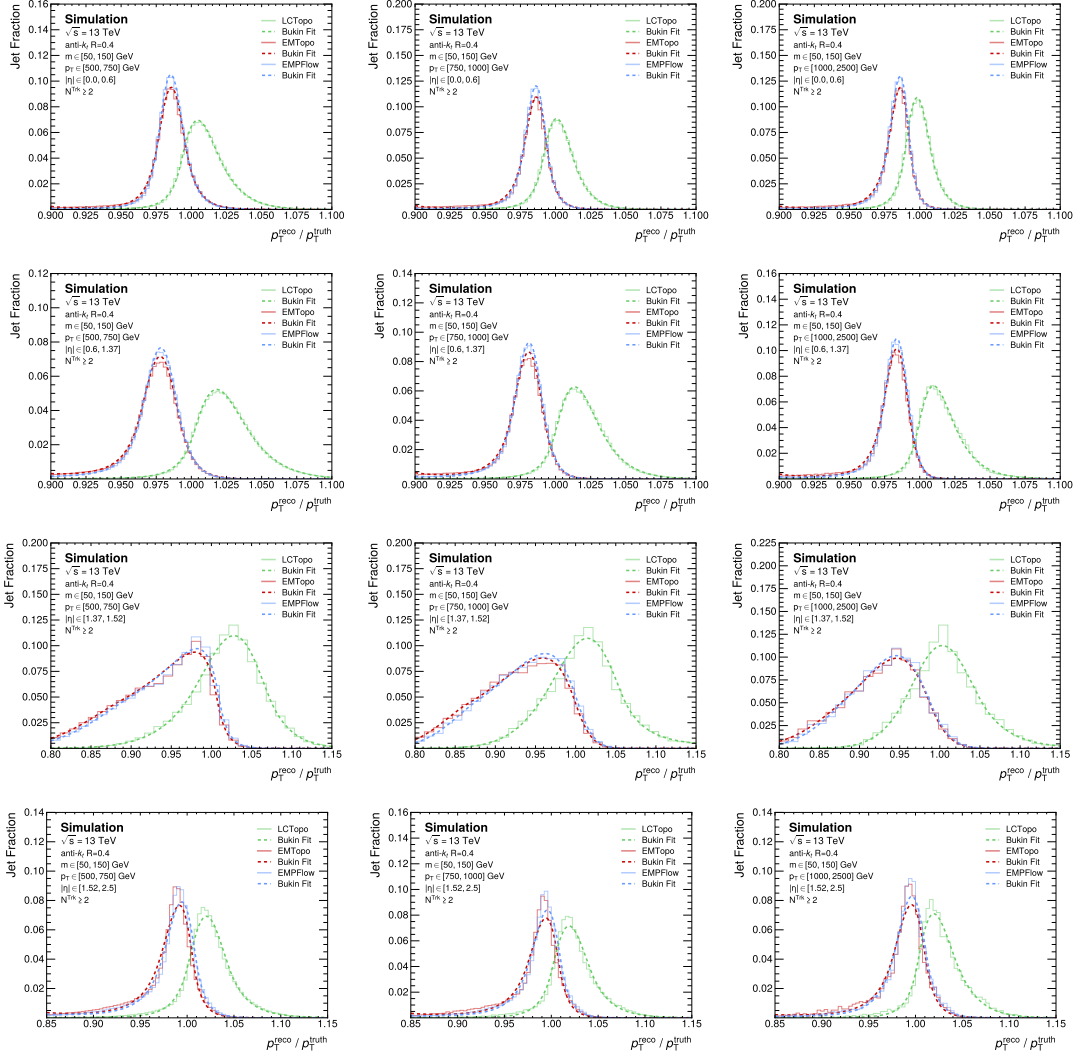


Figure A.7.: Transverse momentum response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with LCTopo (green), EMTopo (red) and EMPFlow (blue) for different p_T and $|\eta|$ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.

A.3. Planar Flow Derivation

$$Pf = 4 \cdot \frac{\det(I_w)}{\text{tr}(I_w)^2}$$

$$= 4 \cdot \frac{|p_1|^2 w_2 \cdot |p_2|^2 w_1}{(|p_1|^2 w_2 + |p_2|^2 w_1)^2} \sin^2(\theta_2 - \theta_1)$$

$$I_w = \frac{1}{M} \left(\sum_i w_i \frac{p_{i,k}}{w_i} \frac{p_{i,l}}{w_i} \right)_{kl}$$

$$= \frac{1}{M} \begin{pmatrix} \frac{p_{1x}^2}{w_1} + \frac{p_{2x}^2}{w_2} & \frac{p_{1x}p_{1y}}{w_1} + \frac{p_{2x}p_{2y}}{w_2} \\ \frac{p_{1x}p_{1y}}{w_1} + \frac{p_{2x}p_{2y}}{w_2} & \frac{p_{1y}^2}{w_1} + \frac{p_{2y}^2}{w_2} \end{pmatrix}$$

$$\det(I_w) = \frac{1}{M^2} \left(\frac{p_{1x}^2}{w_1} + \frac{p_{2x}^2}{w_2} \right) \left(\frac{p_{1y}^2}{w_1} + \frac{p_{2y}^2}{w_2} \right) - \left(\frac{p_{1x}p_{1y}}{w_1} + \frac{p_{2x}p_{2y}}{w_2} \right)^2$$

$$= \frac{1}{M^2} \frac{p_{1x}^2 p_{2y}^2 + p_{2x}^2 p_{1y}^2 - 2p_{1x}p_{1y}p_{2x}p_{2y}}{w_1 w_2}$$

$$= \frac{1}{M^2} \frac{(p_{1x}p_{2y} - p_{2x}p_{1y})^2}{w_1 w_2} \quad (\text{Kartesian} \rightarrow \text{Polar coordinates})$$

$$= \frac{1}{M^2} \frac{|p_1|^2 |p_2|^2 (\cos \theta_1 \sin \theta_2 - \sin \theta_1 \cos \theta_2)^2}{w_1 w_2}$$

$$= \frac{1}{M^2} \frac{|p_1|^2 |p_2|^2}{w_1 w_2} \sin^2(\theta_2 - \theta_1)$$

$$\text{tr}(I_w) = \frac{1}{M} \left(\frac{|p_1|^2}{w_1} + \frac{|p_2|^2}{w_2} \right)$$

$$= \frac{1}{M} \frac{|p_1|^2 w_2 + |p_2|^2 w_1}{w_1 w_2}$$

A.4. Performance on the Training Set

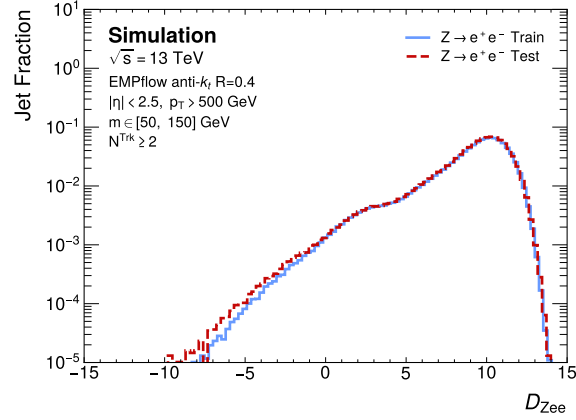


Figure A.8.: Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of $Z \rightarrow e^+e^-$. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.

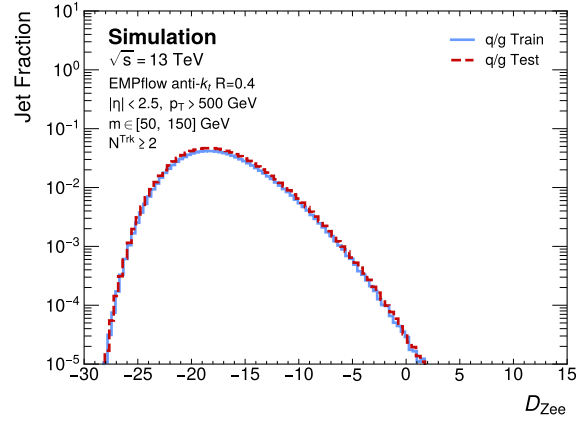


Figure A.9.: Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of q/g -jets. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.

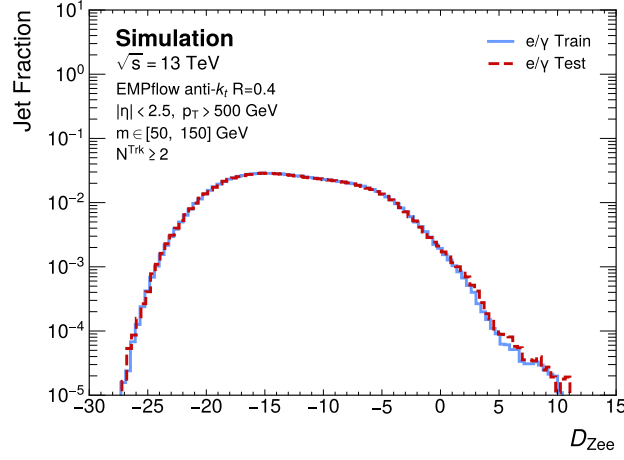


Figure A.10.: Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of e/γ -jets. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.

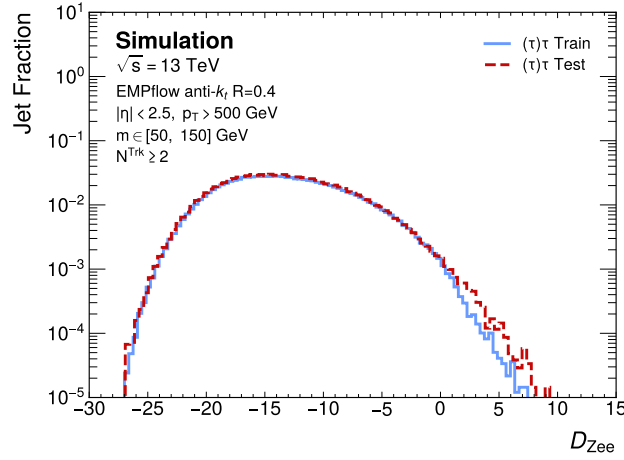


Figure A.11.: Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of $(\tau)\tau$ -jets. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.

List of Figures

2.1.	Particle content of the Standard Model of particle physics. The two types of fermions, quarks (leptons) in green (red), are ordered in the three generations. The gauge bosons are in red, and the Higgs boson in yellow. The physical constants are published by the Particle Data Group [34]. .	5
2.2.	Illustration of the Higgs potential. The SM Higgs potential is not two, but four-dimensional. Selecting a value of the degenerated minimum at the bottom of the potential spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry.	9
2.3.	Pictorial view of the Bridge Method. Theory and data meet at the central pillar. [33]	13
2.4.	Upper panel: Branching Ratios for the two body decays of the neutral vector V^0 for the benchmarks $A_{g_V=1}$ (left) and $B_{g_V=3}$ (right). Lower panel: Total widths corresponding to different values of the coupling g_V in the models A (left) and B (right). [33]	15
3.1.	The CERN accelerator complex. The proton bunches pass the pre-accelerators LINAC2, BOOSTER, PS, and SPS and then reach their final energy in the LHC with the four detectors ALICE, ATLAS, CMS, and LHCb. [56]	17
3.2.	The ATLAS Coordinate system.	18
3.3.	Cut-away view of the ATLAS inner detector. [15]	20
3.4.	A schematic view of the ATLAS calorimeter system. The EM barrel and end-cap calorimeters are shown in green. The EM end-cap has a precision region marked in darker green and an extended region in light green, and the transition from one to the other around $\eta \sim 2.5$ involves a significant change in the material layers. The hadronic Tile calorimeter is shown in light blue while the hadronic end-cap calorimeters based on liquid argon are illustrated in light orange. The forward calorimeters are shown in dark orange. [58]	22
3.5.	Cut-away view of the ATLAS muon system. [15]	23

3.6.	An illustration of an electron trespassing the detector. The red trajectory shows the path of an electron, traversing the tracking system (pixel detectors, silicon-strip detectors, TRT) and then entering the electromagnetic calorimeter. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron. [63]	25
3.7.	Supercluster with a seed electron cluster and a satellite photon cluster. [64]	27
3.8.	The transformed discriminant d'_L for reconstructed electron candidates. The black histogram is for prompt electrons, and the red line is for the sum of backgrounds. [63]	28
3.9.	Representation of iterative jet clustering algorithms as a flowchart. The starting point is the red node corresponding to the set of four-momenta of clusters created by a clustering algorithm.	31
3.10.	Visualisation of the unique signature of bottom quarks in comparison to light quarks. In contrast to light quarks, the bottom quarks form a secondary vertex.	33
3.11.	The light-flavour jet (dashed line) and c-jet rejection factors (solid line) as a function of the b-jet tagging efficiency of the MV2c10 b-tagging algorithm. [79]	34
3.12.	The average energy response as a function of η for different energies. The energy response is shown after origin and pile-up corrections are applied. [83]	36
3.13.	Overview of the cluster and jet collections at ATLAS. Clusters are shown in green, constituent-level jets in blue, and fully calibrated jets in red. The algorithms transforming the different objects are outlined in the grey nodes.	37
4.1.	Visualisation of the underlying concept of supervised learning. The input is fed into the model, which outputs a prediction. Target, loss, and optimiser govern the training procedure; the loss function compares the target and prediction, and the optimiser adjusts the model for an improved prediction with respect to the loss.	39
4.2.	Illustration of an artificial neuron: each component x_i of the input is weighted by a factor w_i . Then, the sum is determined and a bias b is added, before applying the activation function f	40
4.3.	The (a) ReLU and (b) Sigmoid activation function.	42

4.4.	Illustration of a DNN as an ensemble of connected nodes and the analogue matrix representation. The mapping of one layer to the subsequent layer can be thought of as matrix multiplication. The bias vector $\mathbf{b}$ represents the stimulus threshold of each node in a layer. The non-linear activation functions f provides additional complexity.	43
4.5.	Illustration of the loss value in dependence of the model parameters θ_1 and θ_2 . During training, the network parameters θ converge to a set of values, which minimises the loss function L . Each arrow indicates one optimisation step of the DNN's parameters in the direction of the negative gradient as defined in Equation 4.2.	45
4.6.	Visualisation of the loss value in dependence on the model's complexity of a trained model. A small model might not be capable of expressing the dependencies in data. A very large model, however, might learn patterns that are only valid for the training sample (what is called overfitting). The best generalising network has the lowest loss value on the validation sample.	47
5.1.	A pictorial representation of $Z \rightarrow e^+e^-$ decays for different Lorentz boosts. A high Lorentz boost of the Z boson leads to a low angular separation between the e^+e^- pair. The green cone illustrates the small-R jet used to reconstruct the $Z \rightarrow e^+e^-$ decay.	48
5.2.	Efficiency to reconstruct and identify $Z \rightarrow e^+e^-$ decays using standard ATLAS electron reconstruction and identification algorithms as a function of the generator-level Z boson p_T (right) and the angular separation $\Delta R(e^+, e^-)$ between the generator-level objects (right). The efficiencies are presented for the <i>Tight</i> , <i>Medium</i> , and <i>Loose</i> OPs and without applying any identification requirements. Electrons with a transverse energy larger than 25 GeV and a pseudorapidity satisfying $ \eta < 2.47$, excluding the transition regions between calorimeters $ \eta \in [1.37, 1.52]$, are considered.	49
5.3.	Log-transformed likelihood discriminant of the reconstructed electrons from $Z \rightarrow e^+e^-$ decays with a separation between the two electrons of $\Delta R < 0.1$ (blue) and $\Delta R > 0.1$ (red).	50
5.4.	Product of acceptance and efficiency as a function of the pole mass of an exotic heavy vector boson Z' for the $Z' \rightarrow Zh \rightarrow l^+l^-b\bar{b}/c\bar{c}$ decay channel. [20]	51
5.5.	Selection acceptance times efficiency for the $X \rightarrow ZV \rightarrow llqq$ signal events from MC simulations as a function of the resonance mass for ggF/DY (right) and VBF production (left). [19]	52

- 5.6. The acceptance (A) times efficiency (ε) of the HVT W' in the Drell–Yan (left) and VBF (right) signal region for different mass points and for the individual channels, and the sum of all channels. The uncertainty includes both the statistical and experimental systematic components. [21] 52
- 5.7. Efficiency to reconstruct a $Z \rightarrow e^+e^-$ decay via a single anti- k_t -jet with $p_T > 500$ GeV, $|\eta| < 2.5$, $N^{\text{Trk}} \geq 2$, and $50 \text{ GeV} < m < 150 \text{ GeV}$ presented as a function of the generator-level Z boson p_T . The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. 55
- 5.8. Distributions of $m^{\text{reco}}/m^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on particle-flow objects calibrated to the EM scale including JES, with $p_T \in [500, 750]$ GeV and (a) $|\eta| \in [0.0, 0.6]$, (b) $|\eta| \in [0.6, 1.37]$, (c) $|\eta| \in [1.37, 1.52]$ and (d) $|\eta| \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The mass response distributions of further intervals are presented in Figure A.4 in the Appendix. 58
- 5.9. Distributions of $p_T^{\text{reco}}/p_T^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on particle-flow objects calibrated to the EM scale including JES, with $p_T \in [500, 750]$ GeV and (a) $|\eta| \in [0.0, 0.6]$, (b) $|\eta| \in [0.6, 1.37]$, (c) $|\eta| \in [1.37, 1.52]$ and (d) $|\eta| \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The momentum response distributions of further intervals are presented in Figure A.5 in the Appendix. 59
- 5.10. Distributions of $m^{\text{reco}}/m^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on uncorrected topo jets calibrated to the LC scale (green), uncorrected topo jets calibrated to the EM scale (red) and uncorrected particle-flow jets calibrated to the EM scale (blue). Each case is depicted for $p_T \in [1000, 2500]$ GeV and (a) $|\eta| \in [0.0, 0.6]$, (b) $|\eta| \in [0.6, 1.37]$, (c) $|\eta| \in [1.37, 1.52]$ and (d) $|\eta| \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The momentum response distributions of further intervals are presented in Figure A.6 in the Appendix. 60

5.11. Distributions of $p_T^{\text{reco}}/p_T^{\text{truth}}$ presented for $Z \rightarrow e^+e^-$ -jets based on uncorrected topo jets calibrated to the LC scale (green), uncorrected topo jets calibrated to the EM scale (red) and uncorrected particle-flow jets calibrated to the EM scale (blue). Each jet type is depicted for $p_T \in [1000, 2500]$ GeV and (a) $ \eta \in [0.0, 0.6]$, (b) $ \eta \in [0.6, 1.37]$, (c) $ \eta \in [1.37, 1.52]$ and (d) $ \eta \in [1.52, 2.5]$. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. The momentum response distributions of further intervals are presented in Figure A.7 in the Appendix.	61
5.12. Response scale μ of the twelve η , p_T intervals of the uncorrected topo jets calibrated to the LC scale of the mass (left) and transverse momentum (right). The error bars show the resolutions σ . The underlying response distributions are depicted in Figure A.6 and Figure A.7 in the Appendix. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	62
5.13. Response scale μ of the twelve η , p_T intervals of the uncorrected topo jets calibrated to the EM scale of the mass (left) and transverse momentum (right). The error bars show the resolutions σ . The underlying response distributions are depicted in Figure A.6 and Figure A.7 in the Appendix. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	62
5.14. Response scale μ of the twelve η , p_T intervals of the uncorrected particle-flow jets calibrated to the EM scale of the mass (left) and transverse momentum (right). The error bars show the resolutions σ . The underlying response distributions are depicted in Figure A.6 and Figure A.7 in the Appendix. The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	63
5.15. Mass (left) and transverse momentum response distributions (right) for $N^{\text{Vert}} \in [0, 10]$ (red), $N^{\text{Vert}} \in [10, 20]$ (green) and $N^{\text{Vert}} \in [20, 75]$ (blue). The simulated jets are extracted from $Z/\gamma^* + \text{jets}$ events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	63
5.16. The deployment of the DNN is based on Monte Carlo Samples generated with SHERPA 2.2.11. Jets, tracks and clusters are extracted from the MC sample and combined to one object via ΔR -matching. After the preselection of the $Z \rightarrow e^+e^-$ candidate jets, the jets are attributed to the four classes.	66
5.17. Visualisation of the momenta transverse to the jet axis. The jet axis defines the origin of the coordinate system.	68

5.18. Several inputs distributions as defined in Table 5.2. The distributions are shown before each jet type is individually reweighted to have flat p_T and $ \eta $ spectra. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. All distributions are normalised to unity.	71
5.19. Several inputs distributions as defined in Table 5.2. The distributions are shown before each jet type is individually reweighted to have flat p_T and $ \eta $ spectra. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. All distributions are normalised to unity.	72
5.20. Several inputs distributions as defined in Table 5.2. The distributions are shown before each jet type is individually reweighted to have flat p_T and $ \eta $ spectra. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. All distributions are normalised to unity.	73
5.21. Schematic of the DNN architecture containing a 29-dimensional input vector, four hidden layers with 200 nodes each and four output nodes. The output nodes quantify the probabilities for a jet to correspond to the signal ($Z \rightarrow e^+e^-$) or one of the three background classes ($q/g, e/\gamma, (\tau)\tau$).	74
5.22. Confusion matrix of the the signal ($Z \rightarrow e^+e^-$) or one of the three background classes ($q/g, e/\gamma, (\tau)\tau$). The tagger shows a high separation between signal and background and a significant separation between each of the three background classes.	78
5.23. Distribution of the discriminant of the DNN based $Z \rightarrow e^+e^-$ identification algorithm presented separately for signal jets and jets from the three background classes. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	79
5.24. Rejection rates for q/g -jets, e/γ -jets and $(\tau)\tau$ -jets presented as a function of the signal efficiency. Only candidates with $p_T > 500$ GeV, $ \eta < 2.5$, and $50 \text{ GeV} < m_j < 150 \text{ GeV}$ are considered. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	80
5.25. D_{Zee} distributions of $Z \rightarrow e^+e^-$ jets (blue) and q/g jets (red) are shown for $p_T \in [500, 520]$ GeV and $p_T \in [900, 1000]$ GeV. In each p_T interval, an operating corresponding to a signal efficiency of 95 % is determined.	81

5.26. The four operating points correspond to efficiencies $\varepsilon_{Z \rightarrow ee} \in \{0.88, 0.93, 0.95, 0.97\}$. The transparent horizontal bars are the operating points corresponding to the very efficiencies in each p_T interval. An exponential fit is applied to the operating point intervals, as defined in Table 5.6. The Exponential function defines the p_T -dependent operating point (i.e. the upper/lower threshold value of the D_{Zee} distribution in Figure 5.23) . . .	81
5.27. Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the angular separation between the generator-level electron-positron pair. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set. . . .	83
5.28. Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the transverse momentum. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	83
5.29. Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the pseudorapidity. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	84
5.30. Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the number of vertices. The efficiency curves are depicted for the 97%, 95%, and 93% operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	84
5.31. Rejection rates for q/g -jets (top row), e/γ -jets (middle row) and $(\tau)\tau$ -jets (bottom row) presented as a function of the jet p_T (left column) and jet $ \eta $ (right column) and for different operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	85
5.32. Rejection rates for q/g -jets (top row), e/γ -jets (middle row) and $(\tau)\tau$ -jets (bottom row) presented as a function of the number interaction vertices N^{Vert} and for different operating points. Jets are extracted from simulated events produced with the SHERPA 2.2.11 generator and the NNPDF3.0NNLO PDF set.	86

5.33. Efficiency to identify boosted $Z \rightarrow e^+e^-$ decays via the DNN presented as a function of the angular separation between the generator-level electron-positron pair (blue), and the efficiency to identify an electron stemming from a $Z \rightarrow e^+e^-$ decay with the standard method.	89
6.1. An illustration of a hypothetical $Z' \rightarrow Zh \rightarrow e^+e^-b\bar{b}$ decay. Due to the high mass of the Z' , $m_{Z'}$, the Z boson and the Higgs boson decay are emitted back-to-back. The large transverse momenta result in a collimated topology of their decay products. The $Z \rightarrow e^+e^-$ decay is reconstructed as a small-R jet (green), and the Higgs boson decay as a large-R jet (red).	90
6.2. Representative lowest order Feynman diagram for the resonant production of Z' boson via quark–antiquark annihilation.	91
6.3. Product of acceptance and selection efficiency for $Z' \rightarrow Zh \rightarrow \ell^+\ell^-b\bar{b}$ as a function of the resonance mass $m_{Z'}$	95
6.4. Event distributions of m_{Zh} for the $Z \rightarrow e^+e^-$ SR. The vertical axis shows the number of data events divided by the bin width. The signal for the benchmark HVT <i>Model A</i> with $m_{Z'} = 3$ TeV (red) and $m_{Z'} = 5$ TeV (blue), are normalised to 1 fb^{-1} , and are shown as dashed lines.	96
6.5. Upper limits at the 95 % CL on the product of the cross section for $pp \rightarrow Z'$ and the branching fraction to Zh from the 2-lepton (2L) channels, separately for the reconstruction and identification via the standard method (red), and the $Z \rightarrow e^+e^-$ tagger (black). Both signal regions include muons.	99
A.1. The $Z \rightarrow e^+e^-$ reconstruction is significantly reduced if at least two GSF tracks are required.	102
A.2. Mass response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPflow without JES for different p_T and $ \eta $ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.103	
A.3. Transverse momentum response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPflow without JES for different p_T and $ \eta $ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.	104
A.4. Mass response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPflow+JES for different p_T and $ \eta $ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11. . .	105

A.5. Transverse momentum response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with EMPflow+JES for different p_T and $ \eta $ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.	106
A.6. Mass response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with LCTopo (green), EMTopo (red) and EMPflow (blue) for different p_T and $ \eta $ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.	107
A.7. Transverse momentum response distribution of $Z \rightarrow e^+e^-$ jets reconstructed and calibrated with LCTopo (green), EMTopo (red) and EMPflow (blue) for different p_T and $ \eta $ intervals of the jets. The jets are extracted from $Z/\gamma^* + \text{jets}$ events generated with SHERPA2.2.11.	108
A.8. Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of $Z \rightarrow e^+e^-$. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.	110
A.9. Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of q/g -jets. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.	110
A.10. Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of e/γ -jets. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.	111
A.11. Distribution of the discriminant of the neural network based $Z \rightarrow e^+e^-$ identification algorithm presented separately for the train and test sample of $(\tau)\tau$ -jets. Jets are extracted from simulated events produced with the SHERPA2.2.11 generator and the NNPDF3.0NNLO PDF set.	111

List of Tables

5.1.	The Monte Carlo samples used for the deployment of the $Z \rightarrow e^+e^-$ identification algorithm.	53
5.2.	Description of the quantities used to identify highly boosted $Z \rightarrow e^+e^-$ decays via a deep neural network. The track properties and the observables based on the track-cluster matching are considered for both the leading and subleading tracks per jet.	70
5.3.	Boundaries of the p_T -bins used as an input variable of the DNN. During the training of the network, the loss function is weighted such that p_T -bin distribution is flat.	75
5.4.	Boundaries of the $ \eta $ -bins used as an input variable of the DNN. During the training of the network, the loss function is weighted such that $ \eta $ -bin distribution is flat.	75
5.5.	Summary of the network architecture and final choice of hyperparameters used in the training of the boosted $Z \rightarrow e^+e^-$ tagging algorithm. The range of parameter values used in the optimisation of the hyperparameters is shown as well.	76
5.6.	Rejection rates for q/g -jets, e/γ -jets and $(\tau)\tau$ -jets for different p_T dependent operating points of the $Z \rightarrow e^+e^-$ tagging algorithm.	80
5.7.	All input variables are sorted by their mutual information [128] with the DNN output in descending order. The mutual information MI quantifies the relevance of each input distribution X_i for the DNN output Y	88
6.1.	A list of the selection criteria in three signal regions included in the statistical analysis of the Z' model hypotheses.	95

Bibliography

- [1] D. P. Barber et al. "Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies." In: *Phys. Rev. Lett.* 43 (1979), p. 830.
- [2] M. L. e. a. Perl. "Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation." In: *Phys. Rev. Lett.* 35 (22 Dec. 1975), pp. 1489–1492.
- [3] M. e. a. Banner. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\bar{p}p$ collider." In: *Phys. Lett. B* 122 (1983), pp. 476–485.
- [4] G. A. et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c² at the CERN SPS collider." In: *Physics Letters B* 126.5 (1983), pp. 398–410.
- [5] F. e. a. Abe. "Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab." In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631.
- [6] S. e. a. Abachi. "Observation of the Top Quark." In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637.
- [7] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." In: *Phys. Lett. B* 716 (2012), p. 1. arXiv: 1207.7214 [hep-ex].
- [8] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Phys. Lett. B* 716 (2012), p. 30. arXiv: 1207.7235 [hep-ex].
- [9] R. Contino, D. Marzocca, D. Pappadopulo, and R. Rattazzi. "On the effect of resonances in composite Higgs phenomenology." In: *JHEP* 10 (2011), p. 081. arXiv: 1109.1570 [hep-ph].
- [10] K. Agashe, R. Contino, and A. Pomarol. "The Minimal composite Higgs model." In: *Nucl. Phys. B* 719 (2005), pp. 165–187. arXiv: hep-ph/0412089.
- [11] S. Catterall, L. Del Debbio, J. Giedt, and L. Keegan. "MCRG Minimal Walking Technicolor." In: *Phys. Rev. D* 85 (2012), p. 094501. arXiv: 1108.3794 [hep-ph].

- [12] J. R. Andersen et al. “Discovering Technicolor.” In: *Eur. Phys. J. Plus* 126 (2011), p. 81. arXiv: 1104.1255 [hep-ph].
- [13] L. Randall and R. Sundrum. “A Large mass hierarchy from a small extra dimension.” In: *Phys. Rev. Lett.* 83 (1999), pp. 3370–3373. arXiv: hep-ph/9905221.
- [14] L. Evans and P. Bryant. “LHC Machine.” In: *JINST* 3 (2008), S08001.
- [15] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider.” In: *JINST* 3 (2008), S08003.
- [16] CMS Collaboration. “The CMS Experiment at the CERN LHC.” In: *JINST* 3 (2008), S08004.
- [17] ATLAS Collaboration. “Search for resonances decaying into a weak vector boson and a Higgs boson in the fully hadronic final state produced in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Phys. Rev. D* 102 (2020), p. 112008. arXiv: 2007.05293 [hep-ex].
- [18] ATLAS Collaboration. “Search for resonances decaying into photon pairs in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Phys. Lett. B* 822 (2021), p. 136651. arXiv: 2102.13405 [hep-ex].
- [19] ATLAS Collaboration. “Search for heavy diboson resonances in semileptonic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Eur. Phys. J. C* 80 (2020), p. 1165. arXiv: 2004.14636 [hep-ex].
- [20] A. Collaboration. “Search for heavy resonances decaying into a Z or W boson and a Higgs boson in final states with leptons and b -jets in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: (July 2022). arXiv: 2207.00230 [hep-ex].
- [21] A. Collaboration. “Search for resonant $WZ \rightarrow \ell\nu\ell'\ell'$ production in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: (July 2022). arXiv: 2207.03925 [hep-ex].
- [22] CMS Collaboration. “Search for $W\gamma$ resonances in proton–proton collisions at $\sqrt{s} = 13$ TeV using hadronic decays of Lorentz-boosted W bosons.” In: *Phys. Lett. B* 826 (2021), p. 136888. arXiv: 2106.10509 [hep-ex].
- [23] CMS Collaboration. “Search for heavy resonances decaying to WW , WZ , or WH boson pairs in a final state of a lepton and a large jet in proton–proton collisions at $\sqrt{s} = 13$ TeV.” In: *Phys. Rev. D* 105 (2021), p. 032008. arXiv: 2109.06055 [hep-ex].
- [24] CMS Collaboration. “Search for heavy resonances decaying to ZZ or ZW and axion-like particles mediating nonresonant ZZ or ZH production at $\sqrt{s} = 13$ TeV.” In: *JHEP* 04 (2021), p. 087. arXiv: 2111.13669 [hep-ex].

- [25] ATLAS Collaboration. “Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC.” In: *Eur. Phys. J. C* 79 (2019), p. 375. arXiv: 1808.07858 [hep-ex].
- [26] ATLAS Collaboration. *Performance of W/Z taggers using UFO jets in ATLAS*. ATL-PHYS-PUB-2021-029. 2021.
- [27] ATLAS Collaboration. “Identification of boosted Higgs bosons decaying into b -quark pairs with the ATLAS detector at 13 TeV.” In: *Eur. Phys. J. C* 79 (2019), p. 836. arXiv: 1906.11005 [hep-ex].
- [28] ATLAS Collaboration. *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*. ATL-PHYS-PUB-2020-019. 2020.
- [29] CMS Collaboration. “Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques.” In: *JINST* 15 (2020), P06005. arXiv: 2004.08262 [hep-ex].
- [30] ATLAS Collaboration. “Reconstruction and identification of boosted di- τ systems in a search for Higgs boson pairs using 13 TeV proton–proton collision data in ATLAS.” In: *JHEP* 11 (2020), p. 163. arXiv: 2007.14811 [hep-ex].
- [31] 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.” In: *Eur. Phys. J. C* 79 (2019), p. 639. arXiv: 1902.04655 [hep-ex].
- [32] D. Duda and F. J. Kiwit. *Identification of highly boosted $Z \rightarrow e^+e^-$ decays using the ATLAS detector*. Tech. rep. Geneva: CERN, 2022.
- [33] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer. “Heavy vector triplets: bridging theory and data.” In: *Journal of High Energy Physics* 2014.9 (Sept. 2014).
- [34] R.L. Workman et al. (Particle Data Group). “Review of Particle Physics.” In: *Progress of Theoretical and Experimental Physics* 2022.8 (2022), Art.Nr. 083C01.
- [35] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons.” In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323.
- [36] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons.” In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509.
- [37] B. Povh, K. Rith, C. Scholz, and F. Zetsche. *Particles and Nuclei: An Introduction to the Physical Concepts*. Springer, 2004.
- [38] S. L. Glashow. “Partial Symmetries of Weak Interactions.” In: *Nucl. Phys.* 22 (1961), pp. 579–588.

- [39] S. Weinberg. “A Model of Leptons.” In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266.
- [40] A. Salam. “Weak and Electromagnetic Interactions.” In: *Conf. Proc. C* 680519 (1968), pp. 367–377.
- [41] M. Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013.
- [42] W. N. Cottingham and D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. 2nd ed. Cambridge University Press, 2007.
- [43] P. Langacker. “Introduction to the Standard Model and Electroweak Physics.” In: (2009). eprint: arXiv:0901.0241.
- [44] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press, 1995.
- [45] M. D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, Mar. 2014.
- [46] T. Lange. “Applications of Deep Neural Networks in a Top Quark Mass Measurement at the LHC.” Masters’s thesis. Universität Hamburg, 2018.
- [47] F. Sannino and K. Tuominen. “Orientifold theory dynamics and symmetry breaking.” In: *Physical Review D* 71.5 (Mar. 2005).
- [48] R. Foadi, M. T. Frandsen, T. A. Rytlov, and F. Sannino. “Minimal walking technicolor: Setup for collider physics.” In: *Physical Review D* 76.5 (Sept. 2007).
- [49] A. Belyaev, R. Foadi, M. T. Frandsen, M. Järvinen, F. Sannino, and A. Pukhov. “Technicolor walks at the LHC.” In: *Physical Review D* 79.3 (Feb. 2009).
- [50] M. J. Dugan, H. Georgi, and D. B. Kaplan. “Anatomy of a Composite Higgs Model.” In: *Nucl. Phys. B* 254 (1985), pp. 299–326.
- [51] K. Agashe, R. Contino, and A. Pomarol. “The minimal composite Higgs model.” In: *Nuclear Physics B* 719.1-2 (July 2005), pp. 165–187.
- [52] M. Schmaltz and D. Tucker-Smith. “LITTLE HIGGS THEORIES.” In: *Annual Review of Nuclear and Particle Science* 55.1 (Dec. 2005), pp. 229–270.
- [53] A. Hönle. “Searches for new resonances decaying to a Higgs and a vector boson with the ATLAS detector at the LHC.” PhD thesis. Technische Universität München, 2021.
- [54] K. Aamodt et al. “The ALICE experiment at the CERN LHC.” In: *JINST* 3 (2008), S08002.
- [55] A. A. Alves Jr. et al. “The LHCb Detector at the LHC.” In: *JINST* 3 (2008), S08005.

- [56] E. Mobs. “The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019.” In: (July 2019). General Photo.
- [57] ATLAS Collaboration. *ATLAS Inner Detector Alignment Performance with February 2015 Cosmic Ray Data*. ATL-PHYS-PUB-2015-009. 2015.
- [58] A. Collaboration. “Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: (2020). eprint: arXiv: 2007.02645.
- [59] ATLAS Collaboration. *Performance of the muon spectrometer alignment in 2017 and 2018 data*. ATL-MUON-PUB-2021-002. 2021.
- [60] ATLAS Collaboration. “The ATLAS Inner Detector Trigger performance in pp collisions at 13 TeV during LHC Run 2.” In: *Eur. Phys. J. C* 82 (2021), p. 206. arXiv: 2107.02485 [hep-ex].
- [61] ATLAS Collaboration. “Operation of the ATLAS trigger system in Run 2.” In: *JINST* 15 (2020), P10004. arXiv: 2007.12539 [hep-ex].
- [62] K. Brendlinger. “Physics with Electrons in the ATLAS Detector.” Aug. 2018.
- [63] M. Aaboud, G. Aad, B. Abbott, D. C. Abbott, O. Abdinov, B. Abeloos, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, and et al. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV.” In: *The European Physical Journal C* 79.8 (Aug. 2019).
- [64] ATLAS Collaboration. *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*. ATL-PHYS-PUB-2017-022. 2017.
- [65] ATLAS Collaboration. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2.” In: *Eur. Phys. J. C* 77 (2017), p. 673. arXiv: 1704.07983 [hep-ex].
- [66] T. G. Cornelissen et al. “The global χ^2 track fitter in ATLAS.” In: *Journal of Physics: Conference Series* 119.3 (2008), p. 032013.
- [67] ATLAS Collaboration. *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*. ATLAS-CONF-2012-047. 2012.
- [68] W. Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. ATL-LARG-PUB-2008-002. 2008.
- [69] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV.” In: *Eur. Phys. J. C* 81 (2021), p. 578. arXiv: 2012.00578 [hep-ex].

- [70] G. P. Salam. “Towards jetography.” In: *The European Physical Journal C* 67.3-4 (May 2010), pp. 637–686.
- [71] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1.” In: *Eur. Phys. J. C* 77 (2017), p. 490. arXiv: 1603.02934 [hep-ex].
- [72] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector.” In: *Eur. Phys. J. C* 77 (2017), p. 466. arXiv: 1703.10485 [hep-ex].
- [73] R. Kogler, B. Nachman, A. Schmidt, L. Asquith, E. Winkels, M. Campanelli, C. Delitzsch, P. Harris, A. Hinzmann, D. Kar, C. McLean, J. Pilot, Y. Takahashi, N. Tran, C. Vernieri, and M. Vos. “Jet substructure at the Large Hadron Collider.” In: *Reviews of Modern Physics* 91.4 (Dec. 2019).
- [74] M. Cacciari, G. P. Salam, and G. Soyez. “The anti- k_t jet clustering algorithm.” In: *JHEP* 04 (2008), p. 063. arXiv: 0802.1189 [hep-ph].
- [75] ATLAS Collaboration. *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*. ATL-PHYS-PUB-2015-026. 2015.
- [76] M. Tanabashi et al. “Review of Particle Physics.” In: *Phys. Rev. D* 98.3 (2018), p. 030001.
- [77] M. Paganini. *Machine Learning Algorithms for b -Jet Tagging at the ATLAS Experiment*. 2017.
- [78] ATLAS Collaboration. “ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV.” In: *Eur. Phys. J. C* 79 (2019), p. 970. arXiv: 1907.05120 [hep-ex].
- [79] ATLAS Collaboration. “Measurements of b -jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV.” In: *JHEP* 08 (2018), p. 089. arXiv: 1805.01845 [hep-ex].
- [80] ATLAS Collaboration. *Jet global sequential corrections with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 8$ TeV*. ATLAS-CONF-2015-002. 2015.
- [81] A. Collaboration. “Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: (2020). eprint: arXiv: 2007.02645.
- [82] ATLAS Collaboration. “Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector.” In: *Eur. Phys. J. C* 75 (2015), p. 17. arXiv: 1406.0076 [hep-ex].

- [83] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Phys. Rev. D* 96 (2017), p. 072002. arXiv: 1703.09665 [hep-ex].
- [84] ATLAS Collaboration. “Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s} = 8$ TeV proton–proton collisions in the ATLAS detector.” In: *Eur. Phys. J. C* 77 (2017), p. 241. arXiv: 1609.09324 [hep-ex].
- [85] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction in proton–proton collisions at $\sqrt{s} = 7$ TeV with ATLAS.” In: *Eur. Phys. J. C* 72 (2012), p. 1844. arXiv: 1108.5602 [hep-ex].
- [86] ATLAS Collaboration. E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions. ATLAS-CONF-2018-023. 2018.
- [87] T. M. Mitchell. *Machine Learning*. New York: McGraw-Hill, 1997.
- [88] J. Neyman and E. S. Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses.” In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933), pp. 289–337.
- [89] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning Representations by Back-propagating Errors.” In: *Nature* 323.6088 (1986), pp. 533–536.
- [90] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958.
- [91] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.
- [92] E. Bothmann et al. “Event generation with Sherpa 2.2.” In: *SciPost Phys.* 7.3 (2019), p. 034. arXiv: 1905.09127 [hep-ph].
- [93] T. Gleisberg and S. Höche. “Comix, a new matrix element generator.” In: *JHEP* 12 (2008), p. 039. arXiv: 0808.3674 [hep-ph].
- [94] F. Buccioni, J.-N. Lang, J. M. Lindert, P. Maierhöfer, S. Pozzorini, H. Zhang, and M. F. Zoller. “OpenLoops 2.” In: *Eur. Phys. J. C* 79.10 (2019), p. 866. arXiv: 1907.13071 [hep-ph].
- [95] F. Cascioli, P. Maierhöfer, and S. Pozzorini. “Scattering Amplitudes with Open Loops.” In: *Phys. Rev. Lett.* 108 (2012), p. 111601. arXiv: 1111.5206 [hep-ph].
- [96] A. Denner, S. Dittmaier, and L. Hofer. “COLLIER: A fortran-based complex one-loop library in extended regularizations.” In: *Comput. Phys. Commun.* 212 (2017), pp. 220–238. arXiv: 1604.06792 [hep-ph].

- [97] S. Höche, F. Krauss, M. Schönherr, and F. Siegert. “A critical appraisal of NLO+PS matching methods.” In: *JHEP* 09 (2012), p. 049. arXiv: 1111.1220 [hep-ph].
- [98] S. Höche, F. Krauss, M. Schönherr, and F. Siegert. “QCD matrix elements + parton showers. The NLO case.” In: *JHEP* 04 (2013), p. 027. arXiv: 1207.5030 [hep-ph].
- [99] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn. “QCD Matrix Elements + Parton Showers.” In: *JHEP* 11 (2001), p. 063. arXiv: hep-ph/0109231.
- [100] S. Höche, F. Krauss, S. Schumann, and F. Siegert. “QCD matrix elements and truncated showers.” In: *JHEP* 05 (2009), p. 053. arXiv: 0903.1219 [hep-ph].
- [101] S. Schumann and F. Krauss. “A parton shower algorithm based on Catani–Seymour dipole factorisation.” In: *JHEP* 03 (2008), p. 038. arXiv: 0709.1027 [hep-ph].
- [102] ATLAS Collaboration. “The ATLAS Simulation Infrastructure.” In: *Eur. Phys. J. C* 70 (2010), p. 823. arXiv: 1005.4568 [physics.ins-det].
- [103] GEANT4 Collaboration. “GEANT4 a simulation toolkit.” In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303.
- [104] ATLAS Collaboration. *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*. ATL-PHYS-PUB-2016-017. 2016.
- [105] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.” In: *Eur. Phys. J. C* 81 (2020), p. 689. arXiv: 2007.02645 [hep-ex].
- [106] M. Cacciari et al. “FastJet User Manual.” In: *Eur. Phys. J. C* 72 (2012), p. 1896. arXiv: 1111.6097 [hep-ph].
- [107] ATLAS Collaboration. *Constituent-level pile-up mitigation techniques in ATLAS*. ATLAS-CONF-2017-065. 2017.
- [108] M. Cacciari, G. P. Salam, and G. Soyez. “The catchment area of jets.” In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 005–005.
- [109] M. Cacciari and G. P. Salam. “Pileup subtraction using jet areas.” In: *Physics Letters B* 659.1-2 (Jan. 2008), pp. 119–126.
- [110] A. D. Bukin. *Fitting function for asymmetric peaks*. 2007.
- [111] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data.” In: *JINST* 14 (2019), P12006. arXiv: 1908.00005 [hep-ex].

- [112] P. Billoir and S. Qian. “Fast vertex fitting with a local parametrization of tracks.” In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 311.1 (1992), pp. 139–150.
- [113] ATLAS Collaboration. “Performance of jet substructure techniques for large- R jets in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector.” In: *JHEP* 09 (2013), p. 076. arXiv: 1306.4945 [hep-ex].
- [114] A. A. et al. “Boosted objects: a probe of beyond the standard model physics.” In: *The European Physical Journal C* 71.6 (June 2011).
- [115] A. J. Larkoski et al. “Energy Correlation Functions for Jet Substructure.” In: *JHEP* 06 (2013), p. 108. arXiv: 1305.0007 [hep-ph].
- [116] L.G. Almeida et al. “Top Jets at the LHC.” In: *Phys. Rev. D* 79 (2009), p. 074012. arXiv: 0810.0934 [hep-ph].
- [117] L.G. Almeida et al. “Substructure of high- p_T Jets at the LHC.” In: *Phys. Rev. D* 79 (2009), p. 074017. arXiv: 0807.0234 [hep-ph].
- [118] ATLAS Collaboration. *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector*. ATLAS-CONF-2015-029. 2015.
- [119] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2014.
- [120] K. P. Murphy. *Machine learning : a probabilistic perspective*. Cambridge, Mass. [u.a.]: MIT Press, 2013.
- [121] A. Paszke. “Automatic differentiation in PyTorch.” In: (2017).
- [122] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall. *Activation Functions: Comparison of trends in Practice and Research for Deep Learning*. 2018.
- [123] B. Gao and L. Pavel. “On the Properties of the Softmax Function with Application in Game Theory and Reinforcement Learning.” In: (2018). arXiv: 1704.00805v4 [math.OA].
- [124] N. Japkowicz and S. Stephen. “The class imbalance problem: A systematic study.” In: *Intelligent Data Analysis* (2002), 429–449.
- [125] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks.” In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Y. W. Teh and M. Titterton. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, May 2010, pp. 249–256.
- [126] J. Bergstra and Y. Bengio. “Random Search for Hyper-Parameter Optimization.” In: *Journal of Machine Learning Research* 13.10 (2012), pp. 281–305.

- [127] J. Neyman and E. S. Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses.” In: New York, NY: Springer New York, 1992, pp. 73–108.
- [128] R. Shwartz-Ziv and N. Tishby. *Opening the Black Box of Deep Neural Networks via Information*. 2017. eprint: arXiv:1703.00810.
- [129] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. -. Shao, T. Stelzer, P. Torrielli, and M. Zaro. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations.” In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301 [hep-ph].
- [130] T. Sjöstrand et al. “A brief introduction to PYTHIA 8.1.” In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. arXiv: 0710.3820 [hep-ph].
- [131] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. ATL-PHYS-PUB-2014-021. 2014.
- [132] R. D. Ball et al. “Impact of heavy quark masses on parton distributions and LHC phenomenology.” In: *Nucl. Phys. B* 849 (2011), pp. 296–363. arXiv: 1101.1300 [hep-ph].
- [133] R. D. Ball et al. “Parton distributions for the LHC run II.” In: *JHEP* 04 (2015), p. 040. arXiv: 1410.8849 [hep-ph].
- [134] A. Denner et al. “Collier: A fortran-based complex one-loop library in extended regularizations.” In: *Comput. Phys. Commun.* 212 (2017), pp. 220–238. arXiv: 1604.06792 [hep-ph].
- [135] S. Schumann et al. “A Parton shower algorithm based on Catani-Seymour dipole factorisation.” In: *JHEP* 03 (2008), p. 038. arXiv: 0709.1027 [hep-ph].
- [136] S. Höche et al. “A critical appraisal of NLO+PS matching methods.” In: *JHEP* 09 (2012), p. 049. arXiv: 1111.1220 [hep-ph].
- [137] S. Catani et al. “QCD Matrix Elements + Parton Showers.” In: *JHEP* 11 (2001), p. 063. arXiv: hep-ph/0109231.
- [138] S. Höche et al. “QCD matrix elements and truncated showers.” In: *JHEP* 05 (2009), p. 053. arXiv: 0903.1219 [hep-ph].
- [139] C. Anastasiou et al. “High-precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order.” In: *Phys. Rev. D* 69 (2004), p. 094008. arXiv: hep-ph/0312266.
- [140] S. Frixione et al. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction.” In: *JHEP* 09 (2007), p. 126. arXiv: 0707.3088 [hep-ph].

- [141] P. Nason. “A new method for combining NLO QCD with shower Monte Carlo algorithms.” In: *JHEP* 11 (2004), p. 040. arXiv: hep-ph/0409146.
- [142] S. Frixione et al. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method.” In: *JHEP* 11 (2007), p. 070. arXiv: 0709.2092 [hep-ph].
- [143] S. Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX.” In: *JHEP* 06 (2010), p. 043. arXiv: 1002.2581 [hep-ph].
- [144] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. “An introduction to PYTHIA 8.2.” In: *Comput. Phys. Commun.* 191 (2015), p. 159. arXiv: 1410.3012 [hep-ph].
- [145] ATLAS Collaboration. *Studies on top-quark Monte Carlo modelling for Top2016*. ATL-PHYS-PUB-2016-020. 2016.
- [146] M. Beneke, P. Falgari, S. Klein, and C. Schwinn. “Hadronic top-quark pair production with NNLL threshold resummation.” In: *Nucl. Phys. B* 855 (2012), pp. 695–741. arXiv: 1109.1536 [hep-ph].
- [147] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason. “Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation.” In: *Phys. Lett. B* 710 (2012), pp. 612–622. arXiv: 1111.5869 [hep-ph].
- [148] P. Bärnreuther, M. Czakon, and A. Mitov. “Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t} + X$.” In: *Phys. Rev. Lett.* 109 (2012), p. 132001. arXiv: 1204.5201 [hep-ph].
- [149] M. Czakon and A. Mitov. “NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels.” In: *JHEP* 12 (2012), p. 054. arXiv: 1207.0236 [hep-ph].
- [150] M. Czakon and A. Mitov. “NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction.” In: *JHEP* 01 (2013), p. 080. arXiv: 1210.6832 [hep-ph].
- [151] M. Czakon, P. Fiedler, and A. Mitov. “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$.” In: *Phys. Rev. Lett.* 110 (2013), p. 252004. arXiv: 1303.6254 [hep-ph].
- [152] M. Czakon and A. Mitov. “Top++: A program for the calculation of the top-pair cross-section at hadron colliders.” In: *Comput. Phys. Commun.* 185 (2014), p. 2930. arXiv: 1112.5675 [hep-ph].
- [153] ATLAS Collaboration. *Multi-Boson Simulation for 13 TeV ATLAS Analyses*. ATL-PHYS-PUB-2017-005. 2017.

- [154] J. Butterworth et al. “PDF4LHC recommendations for LHC Run II.” In: *J. Phys. G* 43 (2016), p. 023001. arXiv: 1510.03865 [hep-ph].
- [155] ATLAS Collaboration. “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector.” In: *JHEP* 09 (2014), p. 145. arXiv: 1406.3660 [hep-ex].
- [156] R. V. Harlander, A. Kulesza, V. Theeuwes, and T. Zirke. “Soft gluon resummation for gluon-induced Higgs Strahlung.” In: *JHEP* 11 (2014), p. 082. arXiv: 1410.0217 [hep-ph].
- [157] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data.” In: *JINST* 14 (2019), P12006. arXiv: 1908.00005 [hep-ex].
- [158] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. “Asymptotic formulae for likelihood-based tests of new physics.” In: *The European Physical Journal C* 71.2 (Feb. 2011).
- [159] D. Buchin. “Development of an Improved Parametrisation of the ATLAS Detector Performance for Tau Leptons and their Application in Studies of Supersymmetric Models within the pMSSM.” Masters’s thesis. Ludwig-Maximilians-Universität München, 2021.
- [160] S. Grewe. “Search for Charged Higgs Bosons in $H^+ \rightarrow W^+h$ Decays with the ATLAS Detector.” Masters’s thesis. Technische Universität München, 2021.
- [161] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017.
- [162] T. Plehn, A. Butter, B. Dillon, and C. Krause. *Modern Machine Learning for LHC Physicists*. 2022.
- [163] D. Curtin, R. Essig, S. Gori, and J. Shelton. “Illuminating dark photons with high-energy colliders.” In: *Journal of High Energy Physics* 2015.2 (Feb. 2015).
- [164] ATLAS Collaboration. *Performance of mass-decorrelated jet substructure observables for hadronic two-body decay tagging in ATLAS*. ATL-PHYS-PUB-2018-014. 2018.