

UNIVERSITÉ DE GENÈVE

Département de physique nucléaire et corpusculaire

Département de physique nucléaire et corpusculaire

FACULTÉ DES SCIENCES

Prof. Giuseppe Iacobucci

Assist. Prof. Steven Schramm

Vector Boson Tagging in the Context of the Search for New Physics in the Fully Hadronic Final State.

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention physique

par

Sofia Adorni Braccesi Chiassi

d'Italie

Thèse n° 5623

GENÈVE

Atelier d'impression ReproMail

2021



DOCTORAT ÈS SCIENCES, MENTION PHYSIQUE

Thèse de Madame Sofia ADORNI BRACCESI CHIASSI

intitulée :

**«Vector Boson Tagging in the Context of the Search for
New Physics in the Fully Hadronic Final State.»**

La Faculté des sciences, sur le préavis de Monsieur G. IACOBUCCI, professeur ordinaire et directeur de thèse (Département de physique nucléaire et corpusculaire), Monsieur S. SCHRAMM, professeur assistant et codirecteur de thèse (Département de physique nucléaire et corpusculaire), Monsieur T. GOLLING, professeur associé (Département de physique nucléaire et corpusculaire), Monsieur A. COCCARO, docteur (Dipartimento di Fisica dell'università di Genova, Istituto Nazionale Fisica Nucleare Sezione di Genova, Genova, Italia), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 13 décembre 2021

Thèse - 5623 -

Le Doyen

Abstract

This thesis presents the search for narrow resonances decaying into a pair of vector bosons (WW , WZ or ZZ) in proton-proton collision data recorded with the ATLAS detector, as well as the development of identification algorithms for this search.

Di-boson systems are a prime probe for new physics: the Standard Model has precise expectation values for the amplitudes of di-boson interactions and any observed deviation, either in the form of a resonance (predicted by many BSM theories) or in a tail excess, would indicate the presence of new physics.

This thesis focuses on the fully hadronic decay mode of the di-boson system: an excess is searched for in the di-boson mass distribution of events in which the two vector bosons decay into two quarks each. Each vector boson is reconstructed as a single large-radius jet that contains the two quarks produced in the decay of the parent vector boson, and their hadronisation products.

To increase the sensitivity of the search, jets are built from particle-flow inputs, such as Track-CaloClusters (TCCs) and United Flow Objects (UFOs), that combine information from both the tracker and calorimeter portions of the ATLAS detector. Both of these procedures are introduced and motivated.

In the full search performed over data collected between 2015 and 2018 the use of TCCs resulted in a significant improvement in jet substructure performance at high p_T and allowed for the development and optimisation of a new boson identification (tagging) algorithm. The combined use of TCC inputs and the new tagger led to a two times improvement in our ability to identify vector bosons and reject QCD background with respect to the reference configuration.

The search didn't reveal any deviation from the background expectations, and 95% confidence level exclusion limits were set for the existence of new particles predicted by three BSM theories.

The development of UFOs in 2018, resulting in a significant improvement in jet substructure performance over the entire p_T spectrum, justified the exploration of new tagging techniques involving machine learning. The use of a combined approach that exploits both the traditional jet substructure variables and the raw jet constituent information was developed and resulted in up to 4 times improvement in our ability to identify vector bosons and reject QCD background with respect to the TCC jet tagger used in the full search. In addition, this novel approach allows for the identification of the nature of vector bosons (W or Z) while also letting us explore the possibility of identifying their polarisation (longitudinal and transverse).

Dans cette thèse, les données produites lors de la collision entre deux protons, récoltées par l'expérience ATLAS, sont analysées à la recherche d'une résonance se désintégrant en une paire de bosons vecteurs (WW , WZ or ZZ).

Les systèmes composés de deux bosons nous permettent d'enquêter sur l'existence de nouvelle physique: dans le Modèle Standard, l'amplitude des interactions entre deux bosons correspond a une certaine valeur attendue, l'observation d'une différence entre la valeur observée et celle attendue, cela soit-il sous la forme d'une résonance ou d'un excès dans la queue de la distribution, indiquera la présence de nouvelle physique.

Cette thèse se focalise en particulier sur les systèmes à deux bosons dans lesquels chaque boson vecteur se désintègre en deux quarks. Un boson vecteur est reconstruit en tant que jet hadronique à rayon large contenant les deux quarks produits lors de la désintégration du boson vecteur.

Pour accroître la sensibilité de la recherche, les jets hadroniques sont construits à partir d'objets de type "particle-flow", c'est à dire des objets développés pour combiner de manière optimale les informations provenant du trajectographe et celles provenant du calorimètre. Les Track-CaloClusters (TCCs) et les United Flow Objects (UFOs), tous deux objets de type "particle-flow", sont décrits en détail.

Des jets hadroniques construits à partir de TCCs sont utilisés dans l'analyse des données récoltées de 2015 à 2018. L'utilisation de ce nouvel objet a pour résultat une sensible amélioration de la résolution des variables de sous-structure de jets à haute énergie, nous permettant par la suite de développer un nouvel algorithme de classification de bosons vecteurs. L'emploi combiné de ce nouvel algorithme et des TCCs fait doubler, par rapport à la configuration de référence, notre capacité d'identifier les bosons vecteurs et, par conséquent, de rejeter le signal de fond.

Lors de cette analyse, effectuée en 2019, aucune déviation par rapport aux valeurs attendues n'est observée, et des limites d'exclusions à un niveau de confiance de 95% sont fixées sur l'existence de particules prédites par trois modèles BSM.

Le développement des UFOs en 2018, apportant une ultérieure amélioration de la performance des variables de sous-structure pour jets de n'importe quelle énergie, justifie l'exploration de nouvelles techniques d'étiquetage des jets par le biais de l'intelligence artificielle. Une approche combinant les traditionnelles variables de sous-structures avec les données brutes concernant les constituants des jets est conçue et rapporte un gain jusqu'à quatre fois plus élevé par rapport à la configuration employée lors de l'analyse des données de 2019. Cette nouvelle approche permet entre autre d'identifier la nature des bosons vecteurs (W ou Z) et de tester leur état de polarisation (longitudinal et transversal).

Contents

| | |
|---|-----------|
| Introduction | 1 |
| I Theory and the ATLAS Experiment | 3 |
| 1 Theory | 5 |
| 1.1 Standard model of particle physics - Particles and Interactions | 6 |
| 1.1.1 Fermions | 6 |
| 1.1.2 Interactions | 6 |
| 1.2 Quantum Chromodynamics | 8 |
| 1.3 Electroweak theory | 8 |
| 1.3.1 The gauge sector | 9 |
| 1.3.2 The fermion sector | 9 |
| 1.3.3 The Higgs sector | 10 |
| 1.3.4 The Yukawa sector | 12 |
| 1.3.5 Shortcomings | 12 |
| 1.4 Beyond the Standard Model | 14 |
| 1.4.1 Di-boson interactions | 15 |
| 1.4.2 Extended Gauge Symmetry Model | 15 |
| 1.4.3 Radion Model | 15 |
| 1.4.4 Bulk Randall-Sundrum Model | 16 |
| 1.4.5 Heavy Vector Triplet Model | 16 |
| 1.4.6 Previous Diboson resonance searches | 17 |
| 2 ATLAS detector at the Large Hadron Collider | 19 |
| 2.1 Large Hadron Collider | 19 |
| 2.1.1 Accelerator complex | 20 |
| 2.1.2 LHC Timeline | 22 |
| 2.2 ATLAS detector | 22 |
| 2.3 Tracking system | 26 |
| 2.3.1 Silicon detectors | 26 |
| 2.3.2 Pixel detector | 28 |
| 2.3.3 Semiconductor tracker | 28 |
| 2.3.4 Transition radiation tracker: | 28 |
| 2.4 Calorimeter system | 29 |
| 2.4.1 Electromagnetic calorimeter: | 31 |
| 2.4.2 Hadronic calorimeter: | 32 |

| | | |
|------------|--|-----------|
| II | Reconstruction Techniques at the ATLAS Experiment | 35 |
| 3 | Reconstruction at the ATLAS Detector | 37 |
| 3.1 | Topological clusters | 38 |
| 3.2 | Tracks | 39 |
| 3.2.1 | Key definitions | 40 |
| 3.2.2 | Track reconstruction | 41 |
| 3.2.3 | Algorithmic Improvements to track reconstruction for boosted b-hadrons | 42 |
| 3.3 | Particle Flow | 49 |
| 3.3.1 | Track-CaloClusters | 50 |
| 3.3.2 | PFlow | 54 |
| 3.3.3 | United Flow Objects | 55 |
| 4 | Jet reconstruction | 57 |
| 4.1 | Jet reconstruction algorithm | 59 |
| 4.2 | Inputs to jet reconstruction | 62 |
| 4.3 | Jet grooming | 63 |
| 4.3.1 | Trimming | 63 |
| 4.3.2 | SoftDrop | 64 |
| 4.4 | Pile-up mitigation techniques | 65 |
| 4.4.1 | Constituent subtraction | 65 |
| 4.4.2 | SoftKiller | 66 |
| 4.5 | Jet calibration | 66 |
| 4.5.1 | Simulation based calibration | 66 |
| 4.5.2 | <i>In situ</i> energy calibration | 67 |
| 4.5.3 | <i>In situ</i> jet mass calibration | 68 |
| 4.6 | Jet substructure variables | 68 |
| 4.6.1 | Mass | 69 |
| 4.6.2 | Energy correlation variables | 70 |
| 4.6.3 | N-subjettiness | 71 |
| 4.6.4 | Other substructure variables | 71 |
| III | Searching for Resonances with Boson-Tagged Jets | 85 |
| 5 | Search for diboson resonances in hadronic final states | 87 |
| 5.1 | Data and simulated samples | 88 |
| 5.2 | Use of Track-CaloClusters for large-R jet reconstruction | 89 |
| 5.3 | Vector Boson Identification | 90 |
| 5.3.1 | Choice of significance metric | 91 |
| 5.3.2 | Choice of substructure variables and inputs to jet reconstruction | 91 |
| 5.3.3 | Performance evaluation and parametrisation of optimised tagger | 94 |

| | | |
|----------|---|------------|
| 5.3.4 | Additional event selections applied in the analysis | 95 |
| 5.4 | Boson tagging efficiency | 98 |
| 5.5 | Event selection efficiency | 100 |
| 5.6 | Background parametrisation | 102 |
| 5.7 | Systematic uncertainties | 102 |
| 5.8 | Results | 105 |
| 6 | Machine Learning based classification | 107 |
| 6.1 | Machine Learning Basics | 107 |
| 6.1.1 | Datasets | 108 |
| 6.1.2 | Inputs to machine learning algorithms | 109 |
| 6.2 | Multi-Layer Perceptrons and Deep Neural Networks | 111 |
| 6.2.1 | Perceptron | 111 |
| 6.2.2 | Deep Neural Networks | 113 |
| 6.2.3 | Activation functions | 114 |
| 6.2.4 | Convolutional Neural Network | 115 |
| 6.2.5 | Hyperparameters and Regularisation | 117 |
| 6.3 | Edge Convolution | 118 |
| 6.4 | Software | 120 |
| 7 | Improving boson tagging abilities | 121 |
| 7.1 | Polarisation from an experimental point of view | 122 |
| 7.2 | Samples and datasets | 126 |
| 7.3 | DNNs | 127 |
| 7.4 | EdgeConv | 131 |
| 7.5 | Combined | 138 |
| 7.6 | Outlook | 143 |
| | Conclusions | 147 |

INTRODUCTION

The ATLAS detector at the Large Hadron Collider (LHC) was built to investigate the properties of the Standard Model (SM) and with the hope of finding something beyond it. Jet substructure techniques are essential to many searches for new physics and, with the increased energy at the centre of mass of Run 2 to $\sqrt{s} = 13\text{TeV}$, they are gaining more importance.

Jets are collimated sprays of particles originating from the hadronisation process of a quark or gluon. As such, a vector boson (V) decaying hadronically ($V \rightarrow qq$) should be reconstructed as two distinguishable jets. However, the more boosted the parent particle is, the more its decay products will be collimated. For this reason, for example, boosted vector bosons decaying into two quarks can be reconstructed as an individual jet. It is in these cases that substructure techniques become essential: these methods allow us to identify the nature of the parent particle by analysing the structure of the jet. Substructure techniques have mainly been used for the identification of the hadronic decays of heavy vector bosons ($V \rightarrow qq$ where V indicates either a W or a Z boson) [1], or top quarks [2] and to distinguish between light quark and gluon induced jets [3].

Even though the energy deposits of the two quarks, in the case of boosted objects, are contained within a single jet, it is still important that they can be resolved within the jet. With the high energies reached with Run-2 the decay products become so collimated that the calorimeter segmentation becomes a limitation. This is especially relevant for the di-boson analysis for which this thesis was developed. This analysis looks for a specific signature where a new heavy particle (X) decays into two vector bosons that, in turn, decay hadronically ($X \rightarrow VV \rightarrow qqqq$). The more massive this new particle is, the more boosted the W, Z -bosons will be and the more a new approach is needed to overcome the structural limitations of the calorimeter.

The aim of this thesis is to explore the potential improvements in the identification of vector bosons and their properties. These improvements can be brought by innovations in the jet substructure reconstruction by unifying tracker and calorimeter information and/or by introducing new algorithmic techniques in the development of methods to identify the nature of the jets studied.

This thesis is structured as follows. In the first part the theoretical background of the analysis is introduced and a description of the ATLAS detector is given. This first part provides the motivation for a new approach to substructure jet reconstruction and provides a detailed description of the two sub-detectors from which the information was taken, namely the calorimeter and tracker system. In the second part the methodologies for jet reconstruction, calibration and the derivation of substructure properties are described and, in particular, the Track-CaloCluster and United Flow Objects approach are introduced. Finally, in the third part, the development of methods to identify the nature of jets is described in the context of the first pass of the analysis and, after, with the introduction of machine learning techniques.

Part I

Theory and the ATLAS Experiment

1

Theory

Contents

| | | |
|------------|--|-----------|
| 1.1 | Standard model of particle physics - Particles and Interactions | 6 |
| 1.1.1 | Fermions | 6 |
| 1.1.2 | Interactions | 6 |
| 1.2 | Quantum Chromodynamics | 8 |
| 1.3 | Electroweak theory | 8 |
| 1.3.1 | The gauge sector | 9 |
| 1.3.2 | The fermion sector | 9 |
| 1.3.3 | The Higgs sector | 10 |
| 1.3.4 | The Yukawa sector | 12 |
| 1.3.5 | Shortcomings | 12 |
| 1.4 | Beyond the Standard Model | 14 |
| 1.4.1 | Di-boson interactions | 15 |
| 1.4.2 | Extended Gauge Symmetry Model | 15 |
| 1.4.3 | Radion Model | 15 |
| 1.4.4 | Bulk Randall-Sundrum Model | 16 |
| 1.4.5 | Heavy Vector Triplet Model | 16 |
| 1.4.6 | Previous Diboson resonance searches | 17 |

Physics is the science that studies and describes natural phenomena. The laws that govern such phenomena, expressed in the form of equations, are derived through the use of the scientific method that has at its foundation the concepts of observation, measurement and experiment. From this method, mathematical laws and theories can be formulated and then tested both by ensuring they accurately describe what is already known and by their ability to predict new phenomena.

In particular, in the field of fundamental particle physics, the Standard Model of Particle Physics (SM) is the theory that currently best describes the elementary components of matter and their interactions. However, even though its accuracy and predictions have yet to fail, there are some known shortcomings to the theory. Such limitations are the subject of theories of new physics beyond the Standard Model, some of which are the foundation of the searches presented in this thesis.

1.1 Standard model of particle physics - Particles and Interactions

The Standard Model, [4, 5, 6], depicted in figure 1.1, was developed over many years and its current formulation, dating back to 1973-74, is attributed to Glashow, Weinberg and Salam. **Fermions** are the building blocks of matter whereas their interactions are mediated by **bosons**. The Standard Model correctly describes three of the four known forces: indeed, only the electromagnetic, weak and strong forces can be described by relativistic quantum field theories. A proper quantum theory for gravity is, to this day, missing and this force cannot therefore be included in the Standard Model. At the current energy scale, however, the coupling strength of gravity to elementary particles is small and can thus be neglected.

1.1.1 Fermions

Fermions, which are spin one-half particles, form the building blocks of matter. They can further be separated in two sub-categories: **leptons** and **quarks**. The two types of fermions differ in that, while leptons have only electroweak charge, quarks carry, in addition, a strong force charge called **colour** that can take one of three possible values: red, green and blue. The colour analogy comes from the fact that quarks cannot exist freely as stable particles but need to be grouped to form colour neutral hadrons. Hadrons made up of one quark and one antiquark ($q\bar{q}$) are called mesons whereas those made up of three quarks are referred to as baryons ($q_{red}q_{green}q_{blue}$). Leptons and quarks are arranged in three generations⁽¹⁾, or families, of increasing masses.

Fermions also carry an electric charge that is proportional to the electron charge: $Q = \alpha e$, where α is the coefficient and e is the electron charge. For simplicity of notation, the electron charge symbol is dropped and when talking about charge we only refer to the multiplicative factor. Lepton generations consist of a charged, $Q = -1$, particle (e, μ, τ) and a neutral one (ν_e, ν_μ, ν_τ) whereas quark families are made of an up-type particle ($up, charm, top$) with electric charge $Q = +2/3$ and a down-type particle ($down, strange, bottom$), with $Q = -1/3$.

Each fermion has an antiparticle, that is, a particle with the same mass and spin value but opposite electric charges (both electric and colour).

In the Standard Model neutrinos are described as massless, but the discovery of neutrino oscillations in 1998, [9], has proved that these particles actually do have a mass, however small.

1.1.2 Interactions

The Standard Model, from a mathematical point of view, is a gauge theory defined by a Lagrangian that describes the dynamics of the system. The Lagrangian must be renormalisable and observe the symmetries of the system. In particular, the symmetry group of the Standard Model is given by the combination of the $SU(3)$ symmetry group, corresponding to the conservation of the colour charge of the strong force, and the $SU(2)_L \times U(1)_Y$ symmetry group, corresponding to the conservation of the weak isospin (I) and the hyper charge (Y) of the unified electromagnetic and weak forces (electroweak

⁽¹⁾There is no theoretical constraint on the number of generations. The number of light neutrinos can, however, be inferred from its relation to the width of the mass of the Z -boson: $n_\nu < 2.9849 \pm 0.0082$, [7].

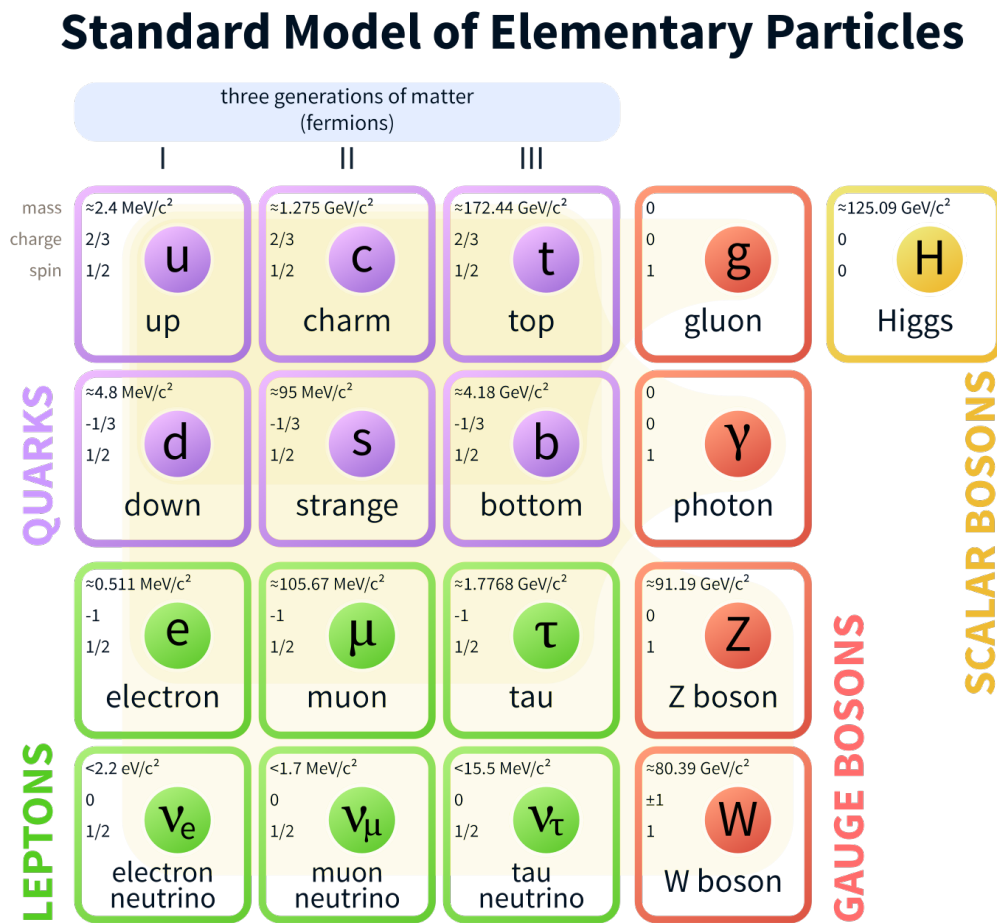


Figure 1.1: The various components of the Standard Model of elementary particles, [8].

interaction). Depending on the symmetry group being unitary (U(N)) or super-unitary (SU(N)), there are N^2 or $N^2 - 1$ generators. The number of generators also indicates the number of mediators of the corresponding force.

The Lagrangian of the Standard Model can therefore be written as the sum of two contributions, the QCD Lagrangian L_{QCD} and the electroweak Lagrangian L_{EW} :

$$L_{SM} = L_{QCD} + L_{EW} \quad (1.1)$$

Let's describe in more detail these two Lagrangians and their consequences.

1.2 Quantum Chromodynamics

The Quantum ChromoDynamics (QCD) Lagrangian involves only quarks q and gluons (bosons, the mediators of the strong force) and is defined as follows:

$$L_{QCD} = -\frac{1}{4} \sum_i F_{\mu\nu}^i F_i^{\mu\nu} + i \sum_r \bar{q}_{r,\alpha} \gamma^\mu D_{\mu,\beta}^\alpha q_r^\beta \quad (1.2)$$

The r term in equation 1.2 indicates the flavour of the quark field and α and β indicate the colour charge of the quarks. γ^μ are the Dirac matrices, which are characterised by a specific anticommutation relation: $\gamma^\mu, \gamma^\nu = 2g^{\mu\nu} I_4$ (I_4 is the four dimensional identity matrix and $g^{\mu\nu}$ is the Minkowski metric). $F_{\mu\nu}^i$ is the gauge invariant field associated to the $SU(3)$ symmetry group [10, 11] and is defined as:

$$F_{\mu\nu}^i = \partial_\mu G_\nu^i - \partial_\nu G_\mu^i - g_F f_{ijk} G_\mu^j G_\nu^k \quad (1.3)$$

The eight G^i ($i \in 1, 2, \dots, 8$) correspond to the eight gluon fields, the g_F constant indicates the value of the strong coupling and the f_{ijk} ($i, j, k \in 1, 2, \dots, 8$) are QCD structure constants. The third term in the above equation represents the gluon self interaction and causes the theory to be non abelian (non commutative). It is interesting to note that the value of the coupling constant g_F is not independent from the energy scale of the interaction process: in the high energy regime the value of the coupling decreases whereas at low energy it increases. This causes two important and opposite phenomena: we observe asymptotic freedom at high energies, meaning that quarks are allowed to propagate freely, and quark confinement at low energies, which is the reason why quarks are not observed as individual particles and are grouped in hadrons. The process of creating stable hadrons when a quark is created in an interaction process is referred to as **hadronisation**.

Finally, the $D_{\mu,\beta}^\alpha$ term is the gauge covariant derivative and is defined as:

$$D_{\mu,\beta}^\alpha = \partial_\mu \mathbb{I}_\beta^\alpha + \frac{i}{2} g_F \sum_i G_\mu^i \lambda_\beta^{i,\alpha} \quad (1.4)$$

where λ^i are the Gell-Mann matrices which generate the group symmetry.

1.3 Electroweak theory

The ElectroWeak (EW) theory accounts for the electromagnetic and weak interactions. It is fully described by the EW Lagrangian that is invariant under the $SU(2) \times U(1)$ symmetry group. The

generators of $SU(2)_L$ are the three Pauli matrices σ_i whereas the generator of $U(1)_Y$ is Y , a 2×2 matrix proportional to the identity. The Lagrangian can be split into four terms each responsible for different phenomena:

$$L_{EW} = L_{gauge} + L_{fermion} + L_\Phi + L_Y \quad (1.5)$$

Written in this form we have that:

- L_{gauge} is responsible for the vector boson dynamics
- The fermion sector $L_{fermion}$ introduces the fermion fields and regulates their interaction with the gauge bosons
- The Higgs sector L_Φ accounts for the existence of the Higgs boson and describes the interaction with gauge bosons and its self-coupling
- L_Y represents the Yukawa sector responsible for fermion masses through the interaction with the Higgs boson field

1.3.1 The gauge sector

The Lagrangian of the gauge sector is given by:

$$L_{gauge} = -\frac{1}{4}W_{\mu\nu}^i W^{\mu\nu i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \quad (1.6)$$

where $W_{\mu\nu}^i$ ($i \in 1, 2, 3$) and $B_{\mu\nu}$ are the gauge invariant field strength associated to the $SU(2)_L$ and $U(1)_Y$ symmetry groups respectively. They are defined as:

$$\begin{aligned} W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\varepsilon^{ijk}W_\mu^j W_\nu^k \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned}$$

In this last equation W_μ^i and B_μ are the actual gauge fields of the electroweak interaction, ε^{ijk} is the Levi-Civita symbol that defines the commutation sign for the canonical quantisation relations that define Lie algebra and g is the coupling constant for the electroweak theory.

It should be noted that if the gauge sector were the only term in the electroweak Lagrangian the theory would predict the existence of four massless bosons [12]. The standard model Lagrangian term that introduces the masses of the vector fields is the Higgs sector.

1.3.2 The fermion sector

The weak interaction violates parity, such a theory is defined as chiral, meaning that it affects right-handed and left-handed fields in different ways. The fermions described in section 1.1.1 are formally represented as a doublet for each generation i : $\psi = (\nu_l, l^-)$ for leptons and $\psi = (u, d)$ for quarks (u are the up-type quarks - u, c, t - and d are the down-type quarks - d, s, b). The two chiral components ψ_L and ψ_R of a generic fermion ψ are:

$$\begin{aligned} \psi_L &= \frac{1}{2}(1 - \gamma^5)\psi \\ \psi_R &= \frac{1}{2}(1 + \gamma^5)\psi \end{aligned}$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ is the fifth gamma matrix.

The interactions of fermions with the gauge bosons are described by the fermion term of the electroweak Lagrangian $L_{fermion}$:

$$L_{fermion} = \sum_{i=1}^3 q_L^i \gamma^\mu D_\mu q_L^i + q_R^i \gamma^\mu D_\mu q_R^i + l_L^i \gamma^\mu D_\mu l_L^i + l_R^i \gamma^\mu D_\mu l_R^i \quad (1.7)$$

where, for simplicity of notation, q denotes a quark doublet and l denotes a lepton doublet and the i indicates the number of generations (3). D_μ is the gauge-covariant derivative:

$$D_\mu = \partial_\mu + 2ig\sigma^i W_\mu^i + ig'Y B_\mu \quad (1.8)$$

with g, g' as the coupling constants and σ^i, Y the generators of $SU(2)_L$ and $U(1)_Y$ respectively.

1.3.3 The Higgs sector

As mentioned in section 1.3.1, the Lagrangian described up until now describes massless particles. Experimental observation of particles with masses renders such a theoretical description incomplete. The Higgs boson [13, 14] corresponds to an isospin doublet complex scalar field $\Phi = (\phi^+, \phi_0)$ (the + and 0 indicate the electric charge of the components) and its existence has been observed in 2012 in proton-proton collisions at the Large Hadron Collider. The Higgs sector of the Standard Model Lagrangian describes the properties of the field and the way it interacts with other particles:

$$L_\Phi = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi^\dagger \Phi) \quad (1.9)$$

$V(\Phi^\dagger \Phi)$ is the potential of the Higgs field and is defined as follows:

$$V(\Phi^\dagger \Phi) = -\mu^2 \Phi^\dagger \Phi + \frac{\lambda}{4} (\Phi^\dagger \Phi)^2 \quad (1.10)$$

As can be seen from the equation two parameters μ and λ characterise the potential. λ is strictly positive since the negative case is considered non physical since such a value doesn't allow the existence of a stable minimum. Two case scenarios present themselves for μ : $\mu^2 > 0$ and $\mu^2 < 0$. In the first case the vacuum expectation value is $\langle 0|\Phi|0 \rangle = 0$ ($|0 \rangle$ corresponds to the ground state), in the other, the minimum of the potential (point of equilibrium) happens when:

$$\Phi^\dagger \Phi = \frac{\mu^2}{2\lambda} \equiv \frac{\nu^2}{2} \quad (1.11)$$

which means that the vacuum expectation value is non-zero: $\langle 0|\Phi|0 \rangle = \frac{\nu}{\sqrt{2}}$.

What are the consequences of this? The non-zero vacuum expectation value means that the equilibrium is degenerate and, thus, the ground state is not symmetric under $SU(2)_L \times U(1)_Y$ transformations. even though Φ is an $SU(2)_L$ doublet, the lowest energy state Φ_0 does not share the symmetry of the group. This phenomenon is referred to as **spontaneous symmetry breaking**.

The physics derived from such a scenario are derived by considering a perturbative expansion of the Lagrangian around the ground state. Three out of the four degrees of freedom of the Higgs field (which

are referred to as Goldstone bosons) disappear after applying a unitary gauge transformation. As a result the Higgs field can be written as:

$$\Phi' = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(\nu + h_0(x)) \end{pmatrix} \quad (1.12)$$

where h_0 is the remaining degree of freedom and corresponds to the particle associated to the Higgs boson field: an electrically neutral real scalar field [10]. The interaction terms between the Higgs field and the gauge fields generates mass terms for W_μ^i and B_μ . These fields however, are not eigenstates of the electromagnetic sub-group ($U(1)_{EM} \subset SU(2)_L \times U(1)_Y$) meaning that the mass matrix obtained from the Higgs mechanism needs to be diagonalised if a massless photon A_μ is to be found. The remaining three bosons are massive and correspond to the W^\pm and Z^0 bosons.

$$\begin{aligned} W_\mu^\pm &= \frac{1}{2}(W_\mu^1 \mp W_\mu^2) \\ Z_\mu^0 &= \cos(\theta_w)W_\mu^3 - \sin(\theta_w)B_\mu \\ A_\mu &= \sin(\theta_w)W_\mu^3 - \cos(\theta_w)B_\mu \end{aligned}$$

where θ_w is the Weinberg weak mixing angle fixed by the coupling constants of the two symmetry groups:

$$\cos(\theta_w) = \frac{g}{\sqrt{g'^2 + g^2}} \quad \text{and} \quad \sin(\theta_w) = \frac{g'}{\sqrt{g'^2 + g^2}} \quad (1.13)$$

The masses of the W and Z vector bosons are computed from interaction and are:

$$m_W = g \frac{\nu}{2} \quad \text{and} \quad m_Z = \frac{\nu}{2} \sqrt{g'^2 + g^2} \quad (1.14)$$

A massive vector field must satisfy the equation of motion that minimise the action. This means that a massive vector boson W_μ must be a solution to the $(\square + m^2)W_\mu + \partial^\mu(\partial_\nu W^\nu)$ equation and, therefore, be of the form:

$$W_\mu = \varepsilon_\mu e^{-ip \cdot x} \quad (1.15)$$

where ε_μ is the polarisation four-vector. For a spin-1 boson travelling along the z -axis, there are three possible polarisations, two transverse ε_-^μ , ε_+^μ and one longitudinal ε_L :

$$\begin{aligned} \varepsilon_-^\mu &= \frac{1}{\sqrt{2}}(0, 1, -i, 0) \\ \varepsilon_L &= \frac{1}{\sqrt{m}}(p_z, 0, 0, E) \\ \varepsilon_+^\mu &= -\frac{1}{\sqrt{2}}(0, 1, i, 0) \end{aligned}$$

The transverse polarisation is composed of right- and left-handed states, with spin either parallel or antiparallel to the momentum vector of the boson. The longitudinal state can only be generated for massive vector bosons. The polarisation rates for vector bosons are interaction dependent and can be calculated by computing the matrix element for the different boson production modes. Physics beyond the standard model can alter the predicted rates, thus making polarisation an important ingredient in the search for new physics.

1.3.4 The Yukawa sector

The Yukawa term of the electroweak Lagrangian couples fermions to the Higgs field. It is defined by:

$$\begin{aligned}
 L_Y &= L_{Y-leptons} + L_{Y-quarks} \\
 L_{Y-leptons} &= - \sum_{i=1}^3 [y_i (l_L^i)^\dagger \Phi l_R^i + h.c.] \\
 L_{Y-quarks} &= - \sum_{i,j=1}^3 [y_i^{d'} (d_L^{i'})^\dagger \Phi d_R^{j'} + y_i^{u'} (u_L^{i'})^\dagger \Phi u_R^{j'} + h.c.]
 \end{aligned}$$

The l and u, d terms are the fermion fields defined in section 1.3.2, the Φ corresponds to the Higgs boson and is responsible for the masses of fermions.

For quarks the mass eigenstates do not match the eigenstates of the gauge boson interactions. As such, the prime in the quarks doublet indicates that the quark in question is the weak eigenstate obtained from the combination of the different quark mass eigenstates following the 3×3 Cabibbo-Kobayashi-Maskawa (CKM) [15, 16] matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (1.16)$$

The fact that the CKM matrix is not diagonal explains the phenomenon of flavour mixing. In the interaction with the Z_0 boson this phenomenon is still not observed (flavour changing flavour currents) whereas it is common for W^\pm bosons (flavour changing neutral currents).

As a summary, the weak force is mediated by the three massive W^\pm and Z bosons, the strong force is mediated by gluons g and the electromagnetic force is mediated by photons γ . All these vector bosons have spin one. The Higgs boson is the mediator of the Higgs scalar field, and, as such, has spin 0.

The interactions between all particles involved in the Standard Model are illustrated in figure 1.2.

1.3.5 Shortcomings

The Standard Model of particle physics has, to date, correctly predicted the mass ratio and existence of the W and Z bosons and the final piece of the puzzle, the Higgs boson, was observed at CERN in 2012 by both the ATLAS and CMS collaborations. These are only some of the numerous successes of the SM, however the theory has some important shortcomings that need to be addressed. Only some of the major ones will be described in the following.

First, it fails to describe the fourth known force: gravity. The gravitational force, accurately described by Einstein's theory of general relativity, is the weakest of all fundamental forces. As such, it is irrelevant at the particle level at the current energy scale, but dominates at large distance scales.

Furthermore, as already mentioned, the observation of neutrino oscillations in 1998 is proof that neutrinos are massive particles. This isn't predicted by the Standard Model and there are two possible ways to explain their masses depending on the intrinsic nature of the neutrino itself. Indeed, neutrinos

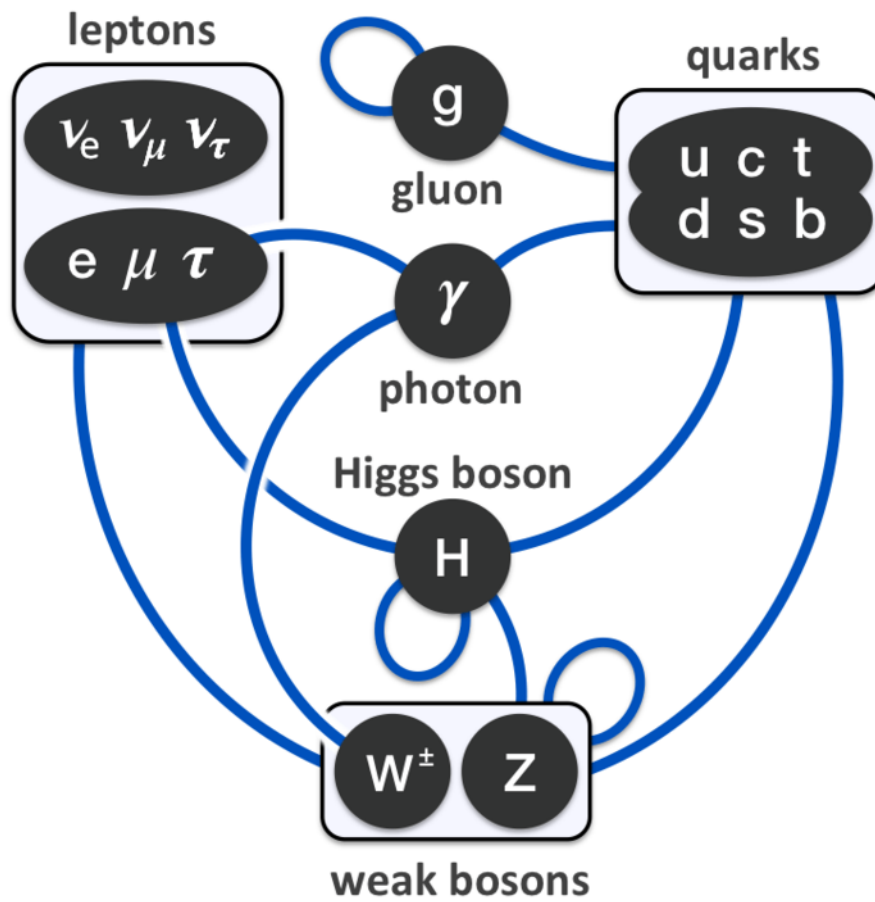


Figure 1.2: The interactions between the particles of the Standard Model, [8].

could have one of two natures: they could be Dirac neutrinos, in this case $\nu \neq \bar{\nu}$, or they could be Majorana neutrinos, in which case the neutrino would be its own anti-particle. This latter case has several implications, ranging from a possible explanation to the matter-antimatter asymmetry to the possibility of the existence of a very heavy neutrino produced by the see-saw mechanism, [17].

The matter-antimatter asymmetry refers to the imbalance in the amount of particles compared to the amount of anti-particles in the observable universe. Some of the asymmetry can be explained by known Standard Model processes resulting in CP violation⁽²⁾, however the amount of such observed processes isn't sufficient to account for the observed asymmetry, [18].

An additional shortcoming of the Standard Model is the hierarchy problem. A hierarchy problem occurs when the expected value of a constant in a Lagrangian (such as the mass) is different from its measured counterpart. For particle physics the problem occurs when trying to explain the discrepancy between the mass of the Higgs boson and the Planck mass. Indeed, the presence of any extension of the Standard Model would induce, in the Higgs propagator, quantum corrections from fermionic loops, of the order of:

$$\Delta m_H^2 = \sum_f -\frac{g_f^2 \Lambda_{UV}^2}{16\pi^2} \quad (1.17)$$

So that the effective mass of the Higgs boson can be written as $m_H^2 = m_{H0}^2 + \Delta m_H^2$, where we refer to m_{H0}^2 as the bare mass. If Λ_{UV} corresponds to the Planck scale (as it would be if the Standard Model were valid up to those energies) a fine-tuning of order 10^{-17} on the bare mass would be needed to obtain the Higgs mass observed. This is not a problem per se but such fine-tuning is rather unnatural. Lastly, beyond the observed asymmetry, around 95 % of the universe is made up of dark energy and dark matter, yet no direct observation of either has been directly made even though their existence and properties can be inferred from indirect measurements. Several hypotheses regarding the nature of the particles composing dark matter and dark energy have been made, but, to date, none has been proven. In the Standard Model, the only candidates for dark matter are neutrinos, but their light mass and weak interaction aren't sufficient to account for such a large percentage of the universe, [17].

1.4 Beyond the Standard Model

All the limitations of the Standard Model, outlined in the previous paragraph, have led physicists to develop new theories to extend its theoretical background. At the moment, no beyond Standard Model theory is able to solve all the shortcomings just mentioned. Between all these theories, it is worth mentioning *SUperSYmmetry* (SUSY): this theory assigns to each particle of the Standard Model a supersymmetric counterpart, [19]. It is able to explain several of the shortcomings of the Standard Model and introduces dark matter particles candidates. However after years of data collection at the LHC, it hasn't been possible to detect any SUSY particle.

Furthermore, the unification of the electromagnetic and weak interaction, has led to the search for a **Grand Unified Theory**, [20, 21], able to describe the three interactions of the Standard Model as

⁽²⁾The violation of the charge conjugation and parity symmetry.

different expressions of a single underlying interaction described by a new symmetry group. [22]. Other theories, relevant to the analysis within which this thesis was developed, are technicolour, [23, 24, 25] and warped extra dimensions, [26, 27, 28].

1.4.1 Di-boson interactions

This thesis, developed in the context of the di-boson analysis [29, 30, 31, 32] focuses on a specific signature: the production of two vector bosons from the decay of a heavy resonance. More specifically, the final state of such events is characterised by the subsequent decay of the vector bosons. The search presented here focuses on the fully hadronic channel, that is, the case in which the vector bosons both decay into pairs of quarks. This channel is the one with the highest branching ratio since $\text{BR}(W \rightarrow q\bar{q}) \approx 68\%$ and $\text{BR}(Z \rightarrow q\bar{q}) \approx 69\%$. However, the search in this channel is limited by the large background due to the overwhelming amount of di-jet events.

All three theories mentioned in the previous section, in fact, predict the existence of a new heavy particle decaying into a vector boson pair (WW , WZ , ZZ). Even though the theoretical model that predicts the heavy resonance does not influence the strategy of the search, which is entirely signature based, benchmark scenarios predicting such particle are necessary for the interpretation of the results. The models used are presented here.

1.4.2 Extended Gauge Symmetry Model

This theory predicts the existence of heavy charged and neutral vector bosons W'^{\pm} and Z' under the assumptions that their couplings to fermions are the same of the known Standard Model weak bosons [33, 34]. The model is explicitly designed to obtain a large production of these events since, otherwise, the observation of the resonance of interest might not be accessible in the search mode. This specific model is used to generate Monte-Carlo samples needed in chapter 5 for experimental optimisation.

1.4.3 Radion Model

The Randall-Sundrum model introduces a solution to the hierarchy problem described above, [26]. The model interprets the four-dimensional space-time as a subset of a higher dimensional bulk with a single warped extra dimension. In the original formulation the Standard Model particles were contained in the usual four dimensions and only the gravitational interaction could propagate through the bulk. This caused gravitational fluctuations in the warped dimension that corresponded to a massless scalar field: the radion. This first model predicted large contributions to flavour changing neutral currents which have been excluded by several observations. Furthermore one other problem with such framework is that the radius of the compactified extra dimension r_c is not determined by the dynamics of the model [35]. One mechanism that solves this problem is the introduction of a bulk scalar radion, produced in gluon-gluon interactions, which interacts on the two ends of the extra dimension [36]. As a result, the radion field of the original formulation of the model acquires a mass term that is on the lower end of the TeV scale. The couplings of the radion to fermions and gauge bosons are proportional to their masses and their squared masses respectively.

If the mass is larger than 1 TeV the main production mode is into a pair of bosons, that is, our signature. The production cross-section times the branching ratio is 6.3 fb for a 2 TeV radion decaying into two W , both of which decay hadronically.

1.4.4 Bulk Randall-Sundrum Model

Another extension of the Randall Sundrum model makes it so that gauge bosons, leptons and quarks can also propagate in the bulk. This extension is referred to as the bulk RS model [26]. In such a scenario the space-time metric must depend on the coordinates of the extra-dimension and therefore becomes:

$$ds^2 = e^{-2kr_c|\phi|} \eta_{\mu\nu} dx^\mu dx^\nu + r_c^2 d\phi^2 \quad (1.18)$$

where $\eta_{\mu\nu}$ is the usual Minkowsky metric in a four dimensional space-time in which the coordinates are four-vectors x^μ , k is the theory energy scale and ϕ is the coordinate of the warped dimension ($0 \leq \phi \leq \pi$). When the warped dimension coordinate is at $\phi = 0$ and $\phi = \pi$ we talk about the Planck and TeV branes respectively. The hierarchy problem between the two branes is solved through the exponential warp factor in equation 1.18.

Indeed, the factor applies to the vacuum expectation value of the Higgs field and, as a consequence, to all masses of the Standard Model. Meaning that if the bare mass of the Higgs boson is at the Planck scale it can be warped down to the TeV scale.

Each Standard Model particle that propagates through the bulk causes Kaluza-Klein (KK) excitations with masses at the TeV scale [37]. The excited KK graviton G_{KK} , produced via gluon-gluon fusion, is expected close to the TeV scale and can decay into: pairs of top quarks, pairs of Higgs bosons, WW and ZZ with significant branching fractions [32]. In particular, for this last decay mode the branching fraction ranges, for WW , between 24% and 20%, and between 12% and 10% for ZZ .

1.4.5 Heavy Vector Triplet Model

The Heavy Vector Triplet (HVT) [38] model Lagrangian introduces a new heavy vector triplet V' (W' and Z') produced via quark-antiquark annihilation. The members of the triplet are degenerate in mass and the couplings to the Standard Model fields are parametrised generally so that the model describes a large class of extensions to the SM. The couplings to the boson and fermion fields are regulated by c_H and c_F respectively. In particular, we have that the coupling to bosons is given by $c_H g_V$ where g_V indicates the strength of the interaction and the coupling to fermions is expressed as $c_F (g^2/g_V^2)$ where g is the $SU(2)_L$ electroweak constant.

Two HVT models, A and B, are used as benchmark scenarios in this thesis.

Model A, which arises from an extension of the SM gauge group, describes a scenario where the three V' are weakly coupled to the Standard Model fields : $g_V = 1$ and $c_F \simeq 1$. In this scenario the branching ratios for the fermion and boson decay modes are comparable. For W' and Z' masses of interest the width of the resonances is narrow. This means that the mass peaks of the V' are observable with a width dominated by experimental resolution

HVT model B, instead, has $g_V = 3$ and $c_F \simeq c_H \simeq 1$ and is representative of composite Higgs models. Since the coupling to fermions is weak, the fermion decay mode of the heavy vector triplet

is suppressed. Meaning that the branching fractions for the di-boson signature of interest are larger than in model A. The width of the resonances, on the other hand, are the same.

1.4.6 Previous Diboson resonance searches

The first ATLAS search in the fully hadronic channel at $\sqrt{s} = 8$ TeV, [29], showed an excess around 2 TeV in the WZ channel with a local significance of 3.4σ and a global significance of 2.5σ . However this result wasn't confirmed by the subsequent search at $\sqrt{s} = 13$ TeV rather exclusion limits were set [30, 31]. In particular, a spin-1 heavy vector triplet predicted by model A of HVT has been excluded for the 1.2 TeV - 3.1 TeV mass range, whereas the existence of a V' from HVT model B has been excluded for masses between 1.2 TeV and 3.5 TeV. The same analysis extended previous lower mass limits for the Kaluza-Klein graviton up to 1.75 TeV. The limits on the production cross sections times branching ratios for these particles have been derived at 95% confidence level.

ATLAS detector at the Large Hadron Collider

Contents

| | | |
|------------|-------------------------------|-----------|
| 2.1 | Large Hadron Collider | 19 |
| 2.1.1 | Accelerator complex | 20 |
| 2.1.2 | LHC Timeline | 22 |
| 2.2 | ATLAS detector | 22 |
| 2.3 | Tracking system | 26 |
| 2.3.1 | Silicon detectors | 26 |
| 2.3.2 | Pixel detector | 28 |
| 2.3.3 | Semiconductor tracker | 28 |
| 2.3.4 | Transition radiation tracker: | 28 |
| 2.4 | Calorimeter system | 29 |
| 2.4.1 | Electromagnetic calorimeter: | 31 |
| 2.4.2 | Hadronic calorimeter: | 32 |

2.1 Large Hadron Collider

The project for the Large Hadron Collider (LHC), [39], [40], [41], was approved in 1994 and its design was conceived to reach an energy at the collision point of up to 14 TeV, thus making it the world's largest and most powerful particle accelerator. The LHC, that consists of a superconducting magnet ring with a 27 km circumference, is only the latest addition to the CERN (Conseil Européen pour la Recherche Nucléaire) accelerator complex.

As its name suggests, the LHC is designed to accelerate hadrons, more specifically, protons and lead ions. Inside the accelerator, two beams of high-energy particles, circulating in opposite directions, are accelerated almost to the speed of light. When they reach their ultimate momentum, they are collided at four different locations around the accelerator ring, corresponding to the positions of the four largest CERN experiments: ATLAS [42], CMS [43], ALICE [44] and LHCb [45].

The LHC accelerator complex is described in the first part of this chapter, followed by the ATLAS experiment and its detectors.

2.1.1 Accelerator complex

The LHC is only the latest addition to CERN's accelerator system. Founded in 1954, CERN has been home to many successful experiments. Several of the already existing accelerators, built for such experiments, have been upgraded and are now part of the LHC injector chain. The entire accelerator complex is depicted in figure 2.1.

The LHC is mainly supplied with protons. Such hadrons are stable and naturally abundant as they can easily be extracted from hydrogen atoms. The choice of protons as colliding particles isn't based solely on those two arguments. Circular accelerators like the LHC lose energy through a process called Bremsstrahlung radiation. This energy loss is proportional to m_{part}^{-4} , where m_{part} is the mass of the colliding particle. Therefore, the energies that can be reached at a circular accelerator with a proton are much larger than the ones reachable, for example, with an electron (also stable and of easy access). However, due to the compound structure of the proton, (uud), the energy of the hard scatter between quarks cannot be accurately predicted and is usually significantly smaller than the (pp) collision point energy.

At the LHC, protons are extracted from a simple bottle of hydrogen, by stripping off the orbital electrons with an electric field. Heavy lead ions are also used in some of the experiments located around the LHC ring to allow the characterisation of quark gluon plasma.

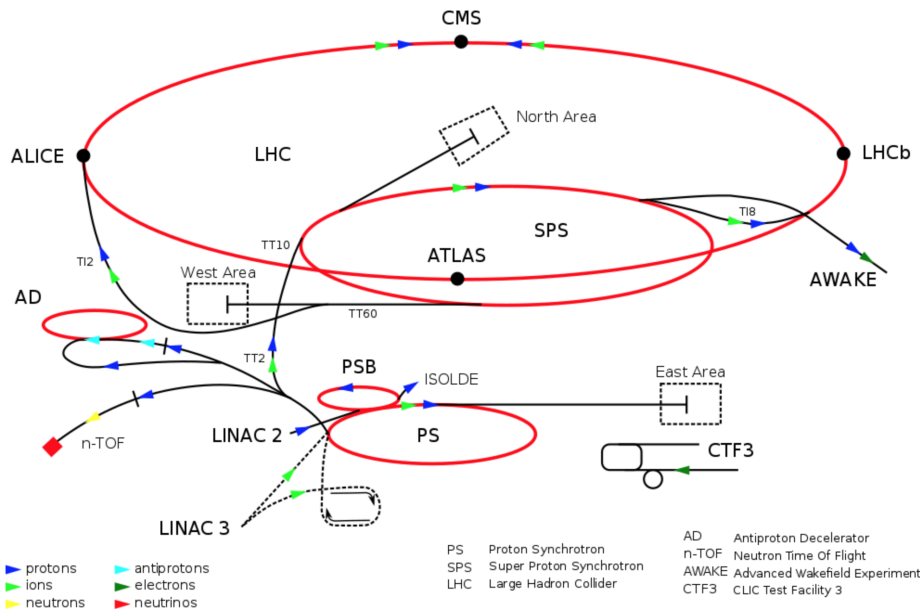


Figure 2.1: CERN accelerator complex [46].

The extracted protons go through the injector chain and are accelerated step by step before reaching the main ring with an energy of 450 GeV. As a first step, the linear accelerator LINAC2 brings the protons to an energy of 50 MeV, then, the Proton Synchrotron Booster accelerates them up to 1.4 GeV before transferring them to the Proton Synchrotron (PS). This circular accelerator takes the proton energy up to 25 GeV, finally, they reach the Super Proton Synchrotron (SPS), the last step before the

LHC ring, which brings them to an energy of 450 GeV. As the name indicates, the injector chain just described injects the accelerated protons into the two beam pipes of the LHC. The beams in question consist of thousands of *bunches*, equally spaced in time and composed of an order of 10^{11} protons. They are accelerated by eight radio frequency cavities that bring them to their colliding energy. This energy varied among the years of existence of the LHC, going from 3.5 TeV in 2010/2011, to the current energy of 6.5 TeV and with the aim of reaching the original design energy of 7 TeV in Run 3 (2021).

The LHC accelerator ring consists of 1232 superconducting dipole magnets with a maximum field strength of 8.33 T that ensure that the beams maintain their circular path.

The LHC was designed with the search for the Higgs boson in mind and a collider whose only feature was to reach high energies wasn't sufficient to make the discovery. Since the Higgs boson production rate dN_{Higgs}/dt is proportional to its cross-section σ_{Higgs} , of the order of picobarn, a very high luminosity, [47, 48], environment was necessary to allow the observation of the new particle in a reasonable amount of time.

$$\frac{dN_{Higgs}}{dt} = \sigma_{Higgs} \cdot L \quad (2.1)$$

where L represents the instantaneous luminosity that is determined by the beam parameters, as seen in equation 2.2.

$$L = \frac{n_b N_1 N_2 f}{A} \quad (2.2)$$

where n_b is the number of bunches in the beam, N_1, N_2 are the number of protons in the two bunches that collide, f is the revolution frequency and A is the overlapping area of the colliding bunches. From the beginning of data-taking in 2010, these parameters have been changed in order to obtain higher luminosity values and the design peak luminosity (of $10^{34} \text{cm}^{-2} \text{s}^{-1}$) has been surpassed as of July 2016. As will be reminded throughout this thesis, such high luminosity conditions creates a challenging environment for the reconstruction of signals of interest. Since the collisions happen between two bunches of protons, several proton-proton collisions can happen at the same time. In addition to the hard scatter collision of interest, there will thus be soft contributions from the simultaneous (pp) collisions⁽¹⁾. This phenomenon is called *in-time pile-up* and is characterised by the number of vertices N_{vtx} reconstructed in the tracker (see section 2.3). At the same time, the small time gap between subsequent bunches makes it possible to have signal residues from the previous collision in the present one: *out-of-time pile-up*. It occurs mainly in parts of the detector that have a readout time longer than the bunch crossing spacing.

Furthermore, to ensure high luminosity, the beams are focused at the interaction points (placed at the location of the four main CERN experiments) by the means of superconducting quadrupole magnets.

⁽¹⁾We refer to particles as hard if they originate from hard scatter collisions, that is from interactions where deep inelastic scattering occurs. As such, these particles carry large amounts of energy. On the opposite side, soft particles are either spectators or debris of the main hard scatter interaction and therefore carry small amounts of energy.

2.1.2 LHC Timeline

The timeline of the LHC is split in Runs which are separated by major upgrades to the experiments and the accelerator complex. The time slots dedicated to such upgrades are the Long Shutdowns (LSs). Further technical maintenance to both the detectors or the LHC can be performed during the Extended Year End Technical Stop (EYETS).

During the first Long Shutdown (LS1), in 2013-2014, the centre-of-mass collision energy was increased from 8 to 13 TeV and that is the energy at which the LHC has been operating during Run 2, from 2015 to 2018. The other changes mainly regard the luminosity: the number of bunches per beam was increased and the bunch spacing reduced. This led the instantaneous luminosity to reach $10^{34} \text{cm}^{-2} \text{s}^{-1}$ and, as a consequence, an integrated luminosity over the full Run2 period of about 150fb^{-1} .

During Long Shutdowns experiments also have the occasion to perform upgrades to their detectors. In particular, ATLAS inserted the IBL layer within the tracking system (see section 2.3) which required the replacement of the beam-pipe and a complete revision of the pixel detector.

Ambitious changes are already planned in the next 10 years to achieve an integrated luminosity of 3000fb^{-1} by 2035, this stage of the LHC is referred to as the High-Luminosity LHC. The complete timeline is shown in figure 2.2.

2.2 ATLAS detector

The ATLAS experiment, [42], is the largest volume detector ever built for a collider. It is a multi-purpose detector designed to measure standard model (SM) parameters but to also explore physics beyond the SM. In order to do this, it is composed of several sub-detectors arranged in concentric layers. The main components are the **Inner Detector**, the **Calorimeter system**, the **magnet system** and the **Muon Spectrometer**. Due to the high radiation environment and the wide variety of particles to be detected, the detector must be designed to be solid and versatile. In addition to the detector system a complex trigger and data acquisition system records the interesting events out of the million happening every second. Both the calorimeter system and the inner detector will be presented in detail, since they are the most relevant to this work, whereas only the main characteristics of the rest of the detector will be described.

All the main components of the experiment can be seen in figure 2.3.

Coordinate system: The origin of the right-handed three-dimensional coordinate system, x, y, z , of the ATLAS detector is located at the nominal collision point. The z axis is aligned with the beam direction, the x -axis points toward the centre of the LHC ring and the y axis is defined to point upwards. An alternative coordinate system η, ϕ, z , is also widely used where ϕ is the azimuthal angle, that is the angle around the beam axis in the xy plane and η is the pseudorapidity. The pseudorapidity, defined as:

$$-\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \quad (2.3)$$

depends on the polar angle θ and is a good approximation of the rapidity, y , for particles of small mass. These coordinates are used because the euclidian distance ΔR , defined in equation 2.4, in the

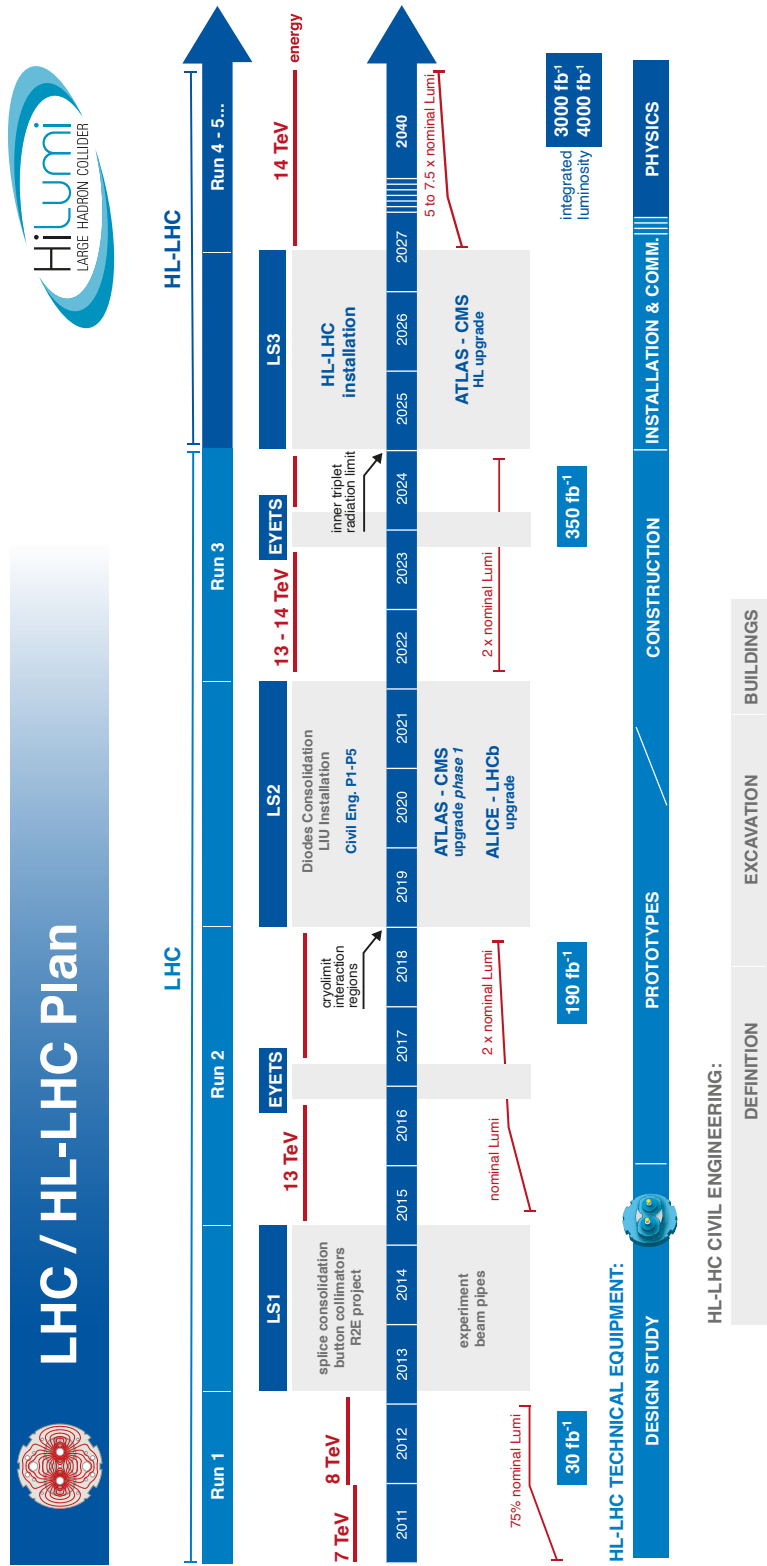


Figure 2.2: Status of the LHC baseline programme. Shown in pictures: the time slots for upgrades of the various experiments (Long Shutdowns - LSs) and for technical maintenance (Extended Year End Technical Stop - EYETS). The picture also shows the increase in the centre-of-mass energy and instantaneous luminosity [49].

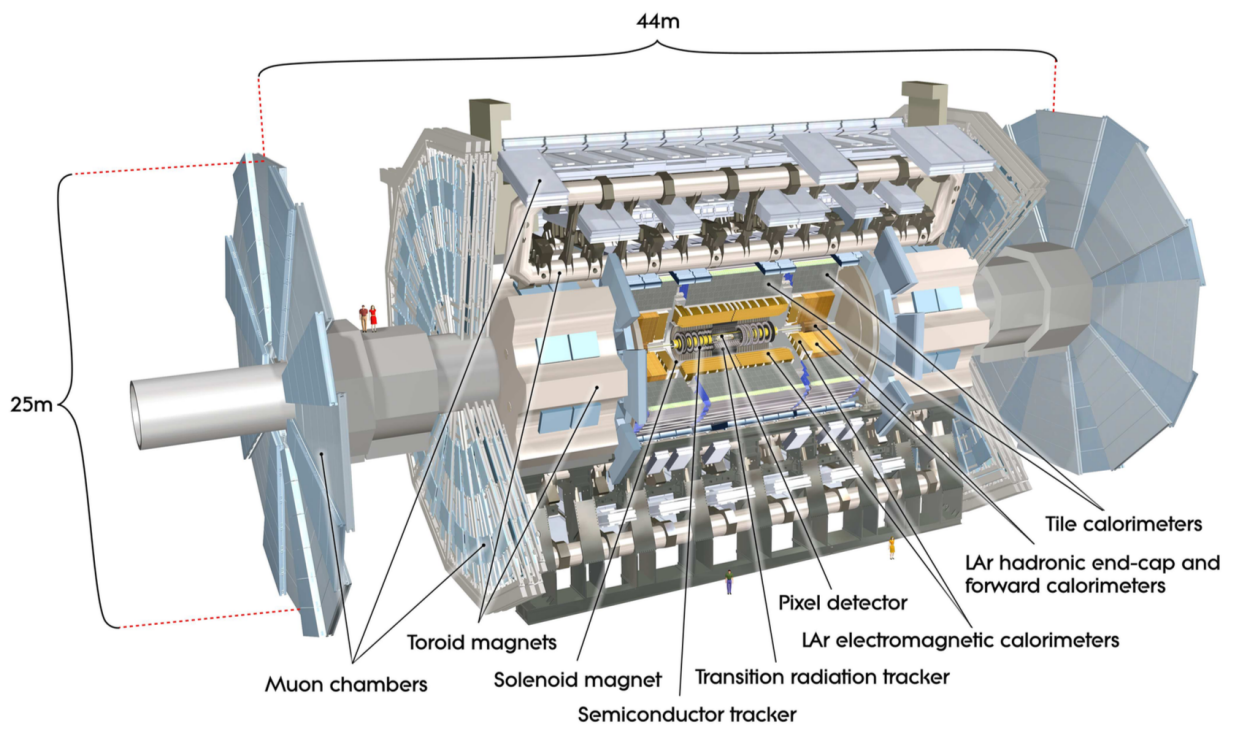


Figure 2.3: Schematic representation of the ATLAS detector, [42].

η, ϕ plane is Lorentz invariant:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (2.4)$$

The transverse momentum p_T is the momentum of the particle projected onto the xy plane. It is common to use p_T measurements in physics analysis at the LHC. Indeed the transverse direction is the only one where the conservation of momentum can be easily observed due to the non-fundamental nature of protons which makes it impossible to know where the primary collision occurs along the longitudinal direction.

Magnet system: The ATLAS magnet system is comprised of the following main parts: the central solenoid, the barrel toroid and two end-cap toroids.

The central solenoid, aligned with the beam axis, generates a 2 T field. This is necessary for the inner detector, discussed in section 2.3, where the bending of a charged particle’s trajectory is used to derive its momentum and charge. The solenoid was designed to minimise the radiative thickness in front of the calorimeter, [42], that is to limit the amount of interactions between incoming particles with the magnet’s material.

The barrel toroid and the two end-cap toroids produce a magnetic field of 1 and 0.5 T for the muon spectrometers in the central and end-cap region respectively [42].

Muon spectrometer: Muons are minimum ionising particles, as such they leave a small amount of energy in the ATLAS detector and can cover large distances. For this reason, the muon system is located at the outermost layer. The muon tracks are bent by the magnetic field generated from the toroidal magnets just described. This magnetic field is mostly orthogonal to the muon trajectory thus minimising the degradation due to multiple scattering. Furthermore, the magnet configuration minimises energy loss in the magnet material by using an “open” structure. The momentum of muon tracks is measured in tracking chambers. More specifically, in the barrel region, corresponding to $|\eta| < 1.4$ the monitored drift tubes are used, whereas, in the transition, $1.4 < |\eta| < 1.6$, and the end-cap, $1.6 < |\eta| < 2.7$, regions cathode strip chambers are used.

Trigger and Data Acquisition The high collision frequencies present at the LHC make it impossible for the detectors to record all of the millions of events that happen every second (design bunch spacing). Therefore, a trigger system is designed to choose only the interesting events out of the very large amount of background. To keep up with the unprecedented luminosity of the LHC, the trigger system is divided in different trigger stages: a low level, purely hard-ware based, *level-1* (L1) trigger, and a *high level trigger*. The latter was, in Run-1, split into two trigger stages: *level-2* (L2), that limits the choice of interesting events by investigating the region of interest (RoI) and the *event filter* (EF) that reconstructs the full event. The two stages were then combined again for Run-2. The entire trigger system brings the collision frequency of up to 40 MHz down to a recording frequency of approximately 1000 events per second.

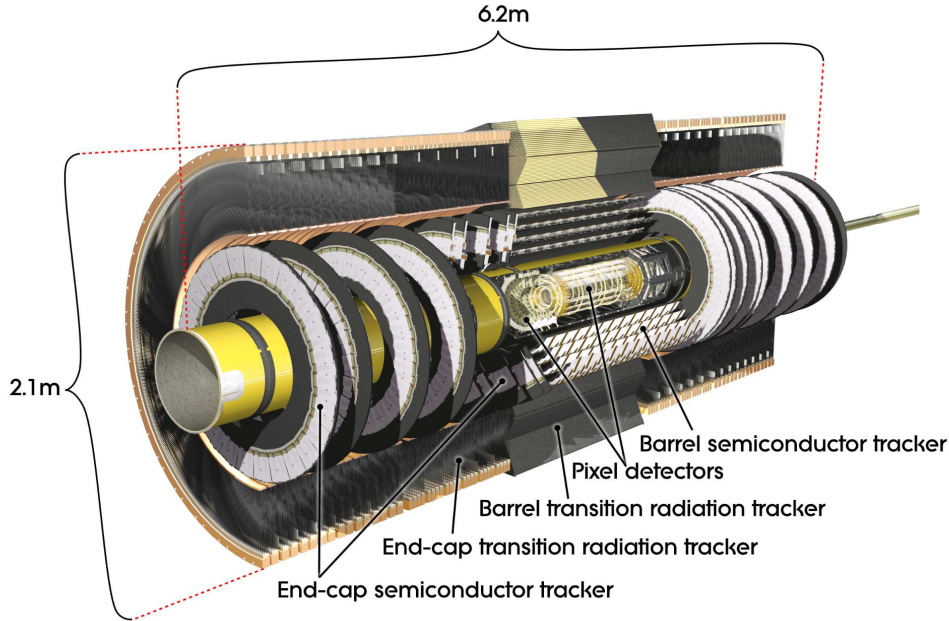


Figure 2.4: Schematic representation of the structure of ATLAS Inner Detector, [42].

2.3 Tracking system

High precision tracking and vertex reconstruction is provided by the **Inner Detector** (ID), [50]. In particular it is designed to provide hermetic and robust pattern recognition for charged particles within the pseudo-rapidity range $|\eta| < 2.5$ and with a transverse momentum as low as 0.4 GeV. The ability to reconstruct vertices is crucial both to distinguish between the hard scatter interactions of interest and pile up, and to identify B-hadrons (see section 3.2.3). As the name suggests, this detector is the innermost layer of the ATLAS detector. As such, it is required to have as little material as possible to, on one side, avoid affecting the energy measurements performed at the calorimeter level and, on the other, to reduce scattering effects within the tracker which would limit our ability to reconstruct a clear track. As shown in figure 2.4, it is composed of three independent sub-detectors. From the interaction point there are subsequently, arranged in order of decreasing granularity, a pixel detector, a silicon strip detector (SCT) and a transition radiation tracker (TRT). Amongst these, the first two provide very fine granularity sensors that are able to resolve the high density environment present at the LHC and therefore allow for precise vertex reconstruction. Both of these tracking systems are solid state detectors using silicon. The TRT is instead used to reconstruct tracks at larger radii, furthermore it is used for particle identification, and, more precisely, to distinguish between electrons and pions.

2.3.1 Silicon detectors

Silicon detectors are widely used in particle physics applications. Indeed, they can provide high granularity and good position resolution. Their principle of operation relies on semiconductors to measure the free charge liberated via ionisation due to the passage of charged particles through the medium. In a solid, the atoms it is composed of are arranged along a crystalline structure (lattice) and tied

together through covalent bonds. All solids are characterised by bands of energy states that are created due to the overlap of electron wave functions. Pauli's principle causes two discrete bands to be created, in particular we distinguish the conduction band, where electrons are free to roam, and the valence band, in which the electrons are tightly bound and remain associated to the lattice atom [51]. In semiconductors the energy gap between the two bands is small which places it, in terms of conductivity, between an insulator and a conductor: the band gap is small enough that valence electrons can jump to the conduction band even at room temperature. When an electron jumps it leaves a hole in the valence band which can then be filled by other electrons: both electrons and holes act as charge carriers. The ability of semiconductors to let electric energy flow through them increases with increasing temperature. Semiconductors can be further split into two categories: intrinsic semiconductors and extrinsic/doped semiconductors. The former correspond to materials that naturally behave as semiconductors, it is however difficult to produce large volumes of such materials since very high purity is required to maintain the conductivity, on top of this they are usually characterised by low concentrations of charge carriers. In doped semiconductors the majority of the charge carriers come from impurities introduced in the lattice sites of the crystal. The doping procedures allow for larger volumes to be created and to have materials with larger densities of charge carriers.

Due to the conductivity features of silicon semiconductors, only 1.14 eV is needed to produce one electron-hole pair, which is very small compared to what would be needed to produce ionisation in a gas detector for instance. Furthermore, the high density of solids compared to gas means that a high number of charge carriers can be created in very small sensors.

A single semiconductor sensor is therefore composed of a bulk material in which the electron-hole pairs are produced, and a readout end able to measure the induced pulse. Single sided silicon detectors are sensors where only one side of the bulk material wafer is segmented into pixels or strips. The bulk material consists in the juxtaposition of n-doped (boron or gallium - meaning that there is an excess of electrons) and p-doped (with arsenic or phosphorus - meaning that there is an excess of holes) silicon to create so called pn junctions. At the interface between the two doped sides the difference in the number of electrons and holes causes the diffusion of surplus charges to the oppositely doped material until thermal equilibrium is reached. The remaining ions create what is called the depletion zone, that is a space free of charge carriers and where an electric field stops further diffusion of carriers from one side to the other. Free charges created in this region from an ionising particle traversing the material can be collected at the junction. The size of the depletion region can be increased using an external reverse bias voltage. Since the concentration of doping atoms is different, using the so called depletion voltage, the bulk material can be fully depleted. This means that the entire sensor volume is sensitive to particles passing through the detector. The electric field created by the bias voltage separates the electrons from the holes and prevents them from recombining. The electrons and holes then drift through the bulk towards the oppositely charged end of the sensor. The readout end which comprises of external electronics is used to measure the current pulse that corresponds to the drift of the charges.

When used close to the interaction point of highly energetic particles, silicon sensors suffer from radiation damage. Since this affects the overall detector performance new technologies and configurations have been developed to mitigate these effects.

2.3.2 Pixel detector

Pixel sensors, which provide the highest granularity, are placed as close as possible to the interaction point. The pixel detector comprises of three layers of pixel sensors in the barrel region and four others on each side of the end-cap region. An additional layer, the insertable b-layer (IBL) was inserted between the nominal interaction point and the first layer in the barrel region during LS1. The distance between the collision and the first layer of the detector was therefore reduced from 5 cm to about 3 cm. Because of this placement the number of pixel layers must be limited to reduce the amount of interaction material close to the collision point. Indeed, a large amount of material in the inner detector can affect negatively the energy measurement in the calorimeter. Furthermore, such proximity results in an amount of radiation much larger than that received by other detectors, and, consequently, a faster degradation of the materials.

When a charged particle passes through a pixel sensor, it creates a number of electron-hole pairs without losing much of its energy. The electrons and holes created drift in opposite directions thanks to an electric field and accumulate on parallel planes. The charge collected on such planes is then transformed into a signal.

The pixel detector therefore provides, with the IBL, four high precision measurements for each charged particle that traverses it. It is also designed to precisely determine the point at which the proton-proton interaction occurred, that is, the vertex. Pixel detector measurements are especially important to identify particles like b -mesons and τ -leptons, as will be discussed in detail in section 3.2.3.

2.3.3 Semiconductor tracker

The semiconductor tracker (SCT) is composed of four layers of silicon microstrip detectors in the barrel region and 9 end-cap disk layers on each side, therefore providing at least four precision points per track. The semiconductor detector covers the intermediate radial range, being placed between the pixel detector and the TRT.

The functioning principle of the silicon modules is the same as the one described for the pixel modules, however, due to the reduced particle density at this distance from the interaction point, silicon strips, with lower granularity, are used instead of pixels. The granularity is still very fine and allows the resolution of tracks which are $200\mu m$ or more apart from each other. This feature is fundamental to provide a good pattern recognition and to be able to distinguish the very large number of tracks that cross the detector at the same time for every collision.

Each silicon module comprises four silicon detectors: on each side, two detectors are bonded to form one longer strip. The two resulting single sided strips are then mounted back to back with an angle of 40 mrad to ensure good resolution especially on the z position of the traversing particle.

2.3.4 Transition radiation tracker:

When a charged particle passes through regions characterised by different dielectric constants, electromagnetic radiation is emitted. This phenomenon is called *transition radiation*. This photon emission can be used for two purposes: not only does it provide information on the trajectory of the traversing particle, the rate of emitted photons varies based on the nature of the particle and can

therefore be used to distinguish between electrons and pions. Due to their small mass, the passing of electrons results in a larger amount of emitted photons when compared to heavier particles (like pions).

The TRT is made of several hundred thousands straws filled with a gas mixture of 70% Xe, 27% CO₂ and 3% O₂. In the barrel, the straws are aligned with the beam axis and can't therefore be used to derive informations about the z position of the crossing particle. In the end-cap they are arranged radially. The resolution on the momentum measurement is poor compared to that of the pixel and silicon sensors (the intrinsic accuracy of the TRT is roughly $130\mu m$), however, the large amount of measurements per track, around 37, compensates for it.

In summary, every track with $|\eta| < 2.5$ is measured with 8 precision space-points (including the b-layer) and approximately 36 TRT straws. The limited dimensions of the inner detector, the strength of the magnetic field that decreases along the z axis and the intrinsic limitations of each sub-detector affect the momentum resolution of the tracking system. The transverse momentum resolution of the inner detector, which is p_T dependent, can therefore be expressed by:

$$\frac{\sigma}{p_T} \approx 3.8 \times 10^{-4} p_T(\text{GeV}) \oplus 0.015 \quad (2.5)$$

2.4 Calorimeter system

A calorimeter is a detector that measures the energy of incoming particles. However, with a high granularity, information on the position of the particles is also obtained. The ATLAS calorimeter system is comprised of two main parts: the **electromagnetic calorimeter** and the **hadronic calorimeter**. The two types of calorimeter allow us to measure all types of particles with the exception of muons, that are minimum ionising particles, and neutrinos, that are characterised by a very weak interaction. All the calorimeters described are sampling calorimeters, that is calorimeters where layers of absorbing material and active material are alternated. The two types of material have different functions: the absorbing material forces the particles to interact and, therefore, produce showers made of secondary, less energetic, particles, while the active material measures the energy deposited by the secondary particles. The energy deposited in the absorbing layer is however missed as it can't be measured. Depending on the nature of the incoming particle, the shower characteristics can be very different. Calorimeters that measure the two types of showers are therefore optimised to account for such differences.

Both the Electromagnetic CALorimeter (ECAL) and the Hadronic CALorimeter (HCAL) are located between the solenoid magnet and the toroid magnets with the ECAL being closer to the interaction point. The two main parts just mentioned are both made of several sub-parts that cover different regions of the detector: the barrel region, the end-cap region and the forward region covering, in total, a pseudo-rapidity range of $|\eta| < 4.9$. The layout of the different parts can be seen in figure 2.5.

The resolution of the a sampling calorimeter is parametrised by the following equation:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (2.6)$$

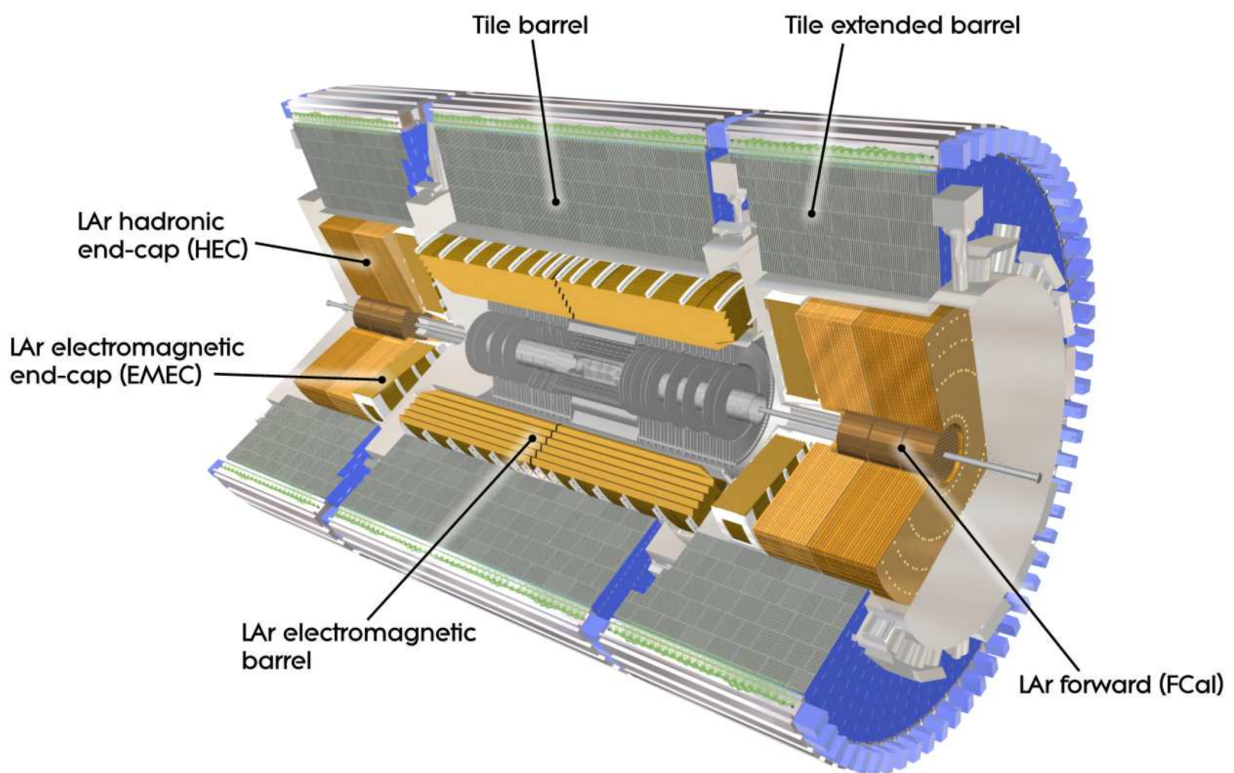


Figure 2.5: Schematic representation of the ATLAS calorimeter, [42].

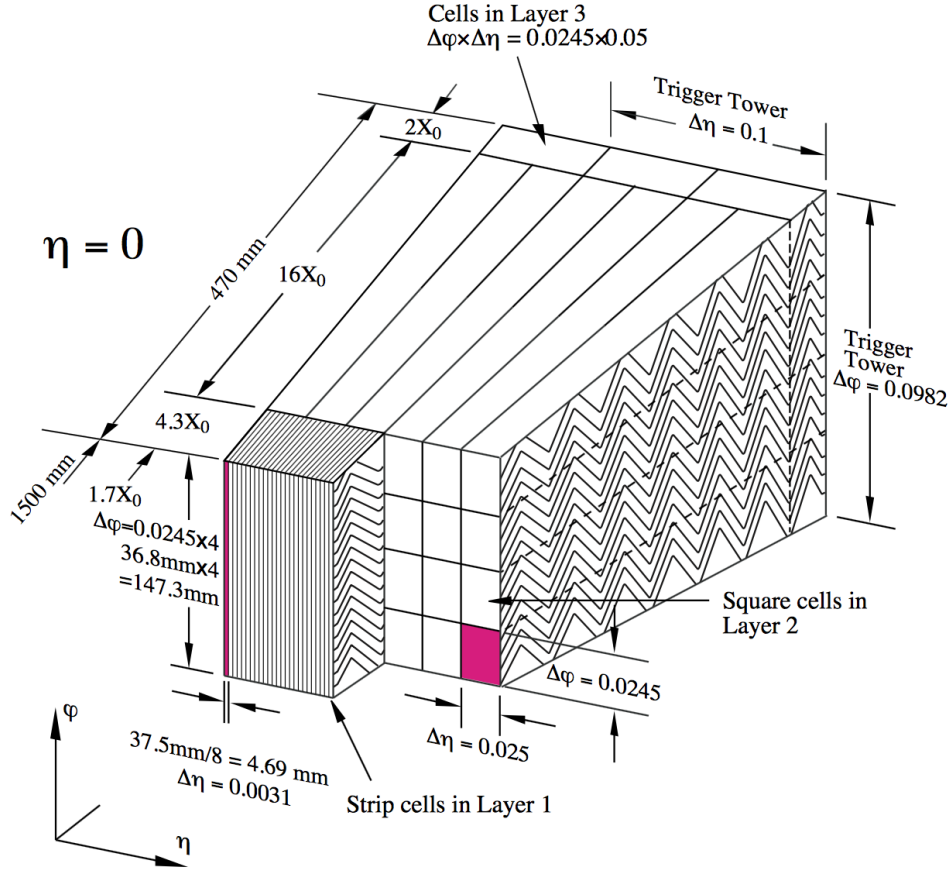


Figure 2.6: Schematic representation of the ATLAS electromagnetic barrel calorimeter, [42].

where the first term, a , is the stochastic term, b is the noise term, due to both pile-up and electronics, and the last term accounts for systematic effects, that is, detector related limitations, [22]. The stochastic term represents the statistical fluctuations of the development of the shower and it is particularly large for sampling calorimeters where the amount of energy deposited in the active material varies from one event to the other (*sampling fluctuations*), [52]. The leading term in this formula depends on the energy level: at low energies, the resolution is mainly limited by the noise while, at high energies, it's the constant c term that affects the performance the most. The parameters for ECAL and HCAL have been estimated to be $b = 10\%$, $c = 0.7\%$ and $b = 50\%$, $c = 3\%$ respectively, [53]. The better resolution of the electromagnetic calorimeter can be explained by its finer granularity ($\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ on average in the barrel region of ECAL, whereas, $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the barrel region of HCAL) and the intrinsic nature of the shower itself, as will be detailed further on.

2.4.1 Electromagnetic calorimeter:

Electromagnetic showers are generated mainly by electrons (e) and photons (γ). Depending on the energy and nature of the incoming particle there are different possible interactions that result in energy loss. At high energies, photons interact primarily via pair production, that is the production

of an electron-positron pair, while electrons emit photons via bremsstrahlung. When the photon energy isn't sufficient to produce the pair, the photo-electric effect becomes the main process for the γ particles, whereas, at similar energy levels, electrons lose energy primarily by ionisation.

Depending on the absorbing material, the radiation length X_0 , defined as the length of material the electron can traverse before losing all but $1/e$ of its energy⁽²⁾, can vary. This value is especially important in the design of the calorimeter since it needs to fully contain the shower in order to measure the energy precisely.

The ATLAS electromagnetic calorimeter (ECAL) is a sampling calorimeter with lead (Pb) as the absorbing material and liquid argon (LAr) as the active material.

The high density of lead makes the value of the radiation length smaller, this is particularly important in the design of the detector as it allows for a more compact design and, therefore, lower costs. Liquid argon, on the other hand, is radiation hard and has an intrinsic linear behaviour, [22], making it a perfect candidate for the active material. The two materials are arranged in an accordion geometry in order to limit cracks (regions where the particles can escape from the detector) and to reach full ϕ coverage. The calorimeter is divided into several region corresponding to different η ranges: the barrel region with $|\eta| < 1.475$, the two end-caps characterised by $1.375 < |\eta| < 3.2$ and finally a layer in the forward calorimeter covers the region $3.1 < |\eta| < 4.9$, [22].

2.4.2 Hadronic calorimeter:

As for electrons and photons, hadrons, both charged and neutral, also produce showers. Hadronic showers, however, are generated by very different processes than electromagnetic ones and are subject to large variations thus making their modelling very complex. This also results in different characteristics. Hadronic showers are generally longer and broader than electromagnetic ones and their size is related to the interaction length⁽³⁾, λ , of the material the particles are traversing.

The hadronic calorimeter measures the energy deposited by hadrons that interact with the absorbing material and then result in said showers. This measurement is of primary importance: for charged hadrons, the energy measured gives complementary information about the particle already detected in the tracker whereas for neutral hadrons it is the only way to measure their energy.

The intrinsic theoretical complexity is only one of the factors limiting the energy resolution of hadronic calorimeters. Indeed, neutral hadrons such as π^0 and η^0 are commonly generated in hadronic processes. Such particles decay into photon-pairs that then generate electromagnetic processes, a large fraction of the deposited energy is therefore of electromagnetic origin. The response of the calorimeter to electromagnetic (e) and hadronic (h) deposits, however, is very different. In ATLAS, all calorimeters are non-compensating, that is, their signal for hadrons is smaller than the one for electrons and photons of the same energy ($e/h > 1$). This happens for several reasons: part of the energy of hadronic showers is missed because it is carried by neutrinos that do not deposit energy in the calorimeter, furthermore, a part of the energy is lost for nuclear processes such as the excitation and fragmentation of atoms, this energy is referred to as *invisible energy*. Applying corrections to

⁽²⁾The length for photons losing energy via pair production is $9/7X_0$ and is therefore related to the radiation length.

⁽³⁾The average distance a hadron can travel in a specific material before an interaction occurs.

the signal locally so that e/h approaches unity on average improves the linearity of the response as well as the energy resolution for jets built from a mix of electromagnetic and hadronic signals. It also improves the reconstruction of full event observables such as $ET_{\text{miss}}^{(4)}$, which combines signals from the whole detector system, [54].

The ATLAS hadronic calorimeter (HCAL) is comprised of several different sub-parts: the tile barrel, the extended tile barrel, two liquid argon, LAr, hadronic end-caps and the forward calorimeter. The tile detectors are sampling calorimeters where iron is used as the absorber material and scintillating tiles serve as the active medium. Both the barrel (covering the range $|\eta| < 1.0$) and extended barrel ($0.8 < |\eta| < 1.7$) are segmented in three layers of different thickness in order to fully contain the shower. The Hadronic End-cap Calorimeters (HEC) use copper plates as absorber and, similarly to ECAL, liquid argon is used as the active material, covering the range $1.5 < |\eta| < 3.2$. With the forward calorimeter, the entire range covered is $|\eta| < 4.9$.

The thickness of the calorimeter, over eleven radiation lengths λ , is chosen with the intention of fully containing showers originating from incoming particles and thus limiting the chance for showers to punch-through the full calorimeter.

In conclusion, it is worth noting the opposite behaviours of the energy/transverse-momentum resolution of the inner detector and calorimeter system: the former has better performance at low transverse momentum and worsens when the energy of the incoming particle increases whereas the opposite is true for the latter. Energy resolution in the calorimeter system improves with the increasing energy, until it reaches its limit, set by the constant term in equation 2.6. These very different characteristics of these two sub-detectors, and the potential to exploit information from both systems, are at the core of the development of the Track Calo-Clusters and United Flow Objects, as will be seen in chapter 3.

⁽⁴⁾ ET_{miss} is the transverse energy missing from the total energy balance of the event and that ideally corresponds to the energy of neutrinos generated in the process.

Part II

Reconstruction Techniques at the ATLAS Experiment

Reconstruction at the ATLAS Detector

Contents

| | | |
|------------|--|-----------|
| 3.1 | Topological clusters | 38 |
| 3.2 | Tracks | 39 |
| 3.2.1 | Key definitions | 40 |
| 3.2.2 | Track reconstruction | 41 |
| 3.2.3 | Algorithmic Improvements to track reconstruction for boosted b-hadrons | 42 |
| 3.3 | Particle Flow | 49 |
| 3.3.1 | Track-CaloClusters | 50 |
| 3.3.2 | PFlow | 54 |
| 3.3.3 | United Flow Objects | 55 |

As seen in the previous chapter, the ATLAS detector is built to interact with the products of high energy collisions, resulting in electrical signals. These signals are digitalised and recorded, and are then processed to become physically interpretable objects. Ultimately, at the analysis level, only a few high level objects, are used to identify the observed events' nature.

Different objects are reconstructed by different parts of the detector: the energy deposits in the calorimeter are grouped in topological clusters (section 3.1), the signals left by charged particles in the tracker are used to reconstruct the particle's trajectory, or track (see section 3.2), the muon detector returns muon trajectory candidates to be matched to the inner detector tracks, etc... Some of these objects naturally carry a physical meaning, for instance each track corresponds to a single charged particle, while others don't. In particular the clusters in the calorimeter need to be combined algorithmically into jets (see next chapter 4) for them to represent an actual hadronic particle⁽¹⁾. Furthermore, objects from different detector parts can be combined amongst themselves to create additional new building blocks that take advantage of the complementarity of the information. That is the concept behind particle flow which we will explore in depth in section 3.3.

⁽¹⁾Jets are collimated sprays of hadrons that originate from quarks and gluons produced in the proton proton collisions. Jets are as close as we can get to a physical single hard quark or gluon.

3.1 Topological clusters

Hadrons emerging from the collisions of the LHC reach the calorimeter and leave a signal and this signal needs to be reconstructed. A topological three-dimensional clustering is performed on the individual calorimeter cells, resulting in topo-clusters [54].

The grouping of the single calorimeter cells is based on the cell signal significance, that is the ratio of the cell energy E_{cell}^{EM} to the cell's average noise $\sigma_{noise,cell}^{EM}$:

$$\zeta_{cell}^{EM} = \frac{E_{cell}^{EM}}{\sigma_{noise,cell}^{EM}} = \frac{E_{cell}^{EM}}{\sqrt{(\sigma_{noise}^{electronic})^2 + (\sigma_{noise}^{pile-up})^2}} \quad (3.1)$$

where the superscript EM indicates that all quantities are measured at the electromagnetic (EM) energy scale⁽²⁾. The cell's average noise corresponds to the squared sum of two noise contributions: the electronic noise $\sigma_{noise}^{electronic}$ and the pile-up noise $\sigma_{noise}^{pile-up}$, that is the noise due to the presence of a number of pp interactions in addition to the main one⁽³⁾.

The topo-cluster formation is a series of *seed and collect* steps governed by three parameters S,N,P.

- If $|\zeta_{cell}^{EM}| > S$ (default value is $S = 4$) the cell is used as seed to the cluster building process. Such seeds each form a *proto-cluster* and are then ordered by decreasing signal significance.
- The neighbouring cells that satisfy $|\zeta_{cell}^{EM}| > N$ (default value is $N = 2$) are collected into the corresponding proto-cluster.
- Finally when there are no more neighbouring cells passing the N threshold, a contour of cells satisfying $|\zeta_{cell}^{EM}| > P$ (default value is $P = 0$) is added.

This procedure is iteratively applied until each seed cell has been clustered. The choice of $P = 0$ means that any cell neighbouring a cell with signal significance larger than $N = 2$ is automatically collected in the proto-cluster, this allows the retention of cells with energies close to the noise levels without degrading the noise suppression feature of the clustering algorithm, [54]. It is also important to note that the thresholds are applied to the absolute value of the cell significance, thus cells with negative signal significance are taken into account. Random positive and negative fluctuations caused by pile-up can therefore cancel each other out. However, only clusters with positive net energy $E_{cluster}$ are considered in jet reconstruction. The proto-clusters just described can be too large to provide a good resolution on the energy flow from the particles of the event. Large clusters with two or more local maxima (defined by cells belonging to the electromagnetic sampling layers - section 2.4 - with $E_{cell}^{EM} > 500$ MeV) are therefore split. Two or more signal maxima can share the same neighbouring cells. The algorithm is built so that each cell is shared once at most: the neighbouring cell is assigned to the two highest energy clusters resulting from the splitting of the original topo-cluster is was associated to. The energy sharing of such a cell between the two clusters is governed by a geometrical weighting

⁽²⁾This scale correctly reconstructs the energy deposits of electrons and photons but does not account for the non compensating character of the calorimeter

⁽³⁾The number of additional pp interactions, μ , varies per year and is based on the expected average $\langle \mu \rangle$ for the run year [54]

process.

Now that the topo-clusters have been defined it is important to describe what exactly they represent. The original motivation of such clustering process is to reconstruct single-particle showers with the highest possible precision of energy and shape. A single topo-cluster is therefore expected for each particle. If this is possible for electromagnetic showers, characterised by their high density and small fluctuations in longitudinal extensions, it is most definitely not conceivable for hadronic showers, which are subject to large fluctuations both in shape and compactness [54]. Hence, topo-clusters do not, individually, represent physical objects. As there is no physically meaningful mass without a specific and valid particle hypothesis for the origin of the signal, [54], topo-clusters are interpreted as massless pseudo-particles, $m_{cluster} \equiv 0$. The remaining kinematical variables that define the cluster four-vector $(E_{cluster}, \eta_{cluster}, \phi_{cluster})$ are derived from the recombination of cell energies and directions.

The topo-cluster size is calculated with respect to the shower axis and the centre of gravity and can be estimated as:

$$\sigma_\eta \approx \sigma_\phi \approx \text{atan}\left(\frac{\sqrt{\langle r^2 \rangle}}{|\vec{c}|}\right) \times \cosh(\eta) \quad (3.2)$$

where $\langle r^2 \rangle$ is the average radial distance to shower axis and \vec{c} is the centre of gravity, for more detail on how to derive these quantities see [54]. The topo-clusters described until now are built at the electromagnetic energy scale. To account for the non-compensating character of the calorimeter, they are then calibrated with the *Local Hadronic Cell Weighting* (LCW) procedure. This method assigns to each topo-cluster a probability $0 \leq P_{cluster}^{EM} \leq 1$ of it being generated from an electromagnetic shower (a topo-cluster generated from a hadronic shower will therefore have $P_{cluster}^{EM} \approx 0$). Depending on the value of this probability the clusters are calibrated to account, in different measures, for the non compensating character of the calorimeter (especially relevant for hadronic topo-clusters) and for the energy losses introduced by clustering and inactive detector material (especially relevant for electromagnetic topo-clusters).

3.2 Tracks

Over the course of history, the developments of silicon based technologies have resulted in smaller and cheaper components thus allowing particle detectors to grow in size and granularity. This has significantly improved our ability to reconstruct the trajectories of charged particles in a magnetic field, also referred to as **tracks**. From these we can derive:

- the track curvature, and, as a consequence, the particle's transverse momentum, p_T .
- the primary and secondary vertices.
- separation amongst nearby particles.

Due to the information we can extract from them, tracks are used in several aspects of jet reconstruction. They are also a crucial instrument in the identification of long-lived b-hadrons and hadronic tau lepton decays.

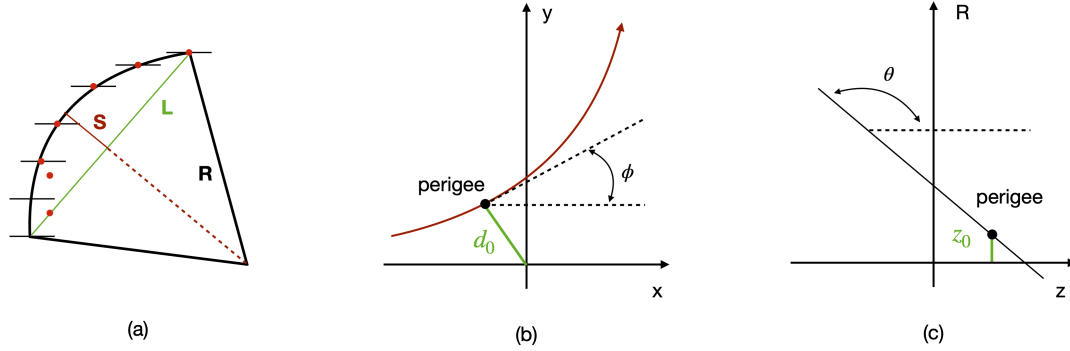


Figure 3.1: (a) The sagitta S and how it relates to the chord length L and the radius R . In a uniform magnetic field, perpendicular to the plane, and without energy loss the trajectory is circular. (b) Graphic representation of the transverse impact parameter d_0 and the perigee point. (c) Graphic representation of the longitudinal impact parameter z_0 , the nominal beam line is aligned with the z axis, as such the perigee point is determined in the transverse plane.

3.2.1 Key definitions

Before going through the details of the complex chain of algorithms that is needed to get to the final object that is a track, let's define how the quantities described above are derived and which quantities are used to measure how well a track is reconstructed.

The trajectory's transverse momentum, p_T , is reconstructed from the sagitta (see figure 3.1), i.e how bent the track is. Indeed, in a magnetic field the curvature of the track relates to its momentum and to the magnitude of said field. The latter being fixed in the detector environment, we can determine the transverse momentum. The transverse momentum resolution is none other than the ratio of the uncertainty on the measurement and the measurement itself: $\frac{\sigma(p_T)}{p_T}$. The primary vertex is defined as the position at which the hard (primary) pp interaction occurs. If particles with macroscopic lifetimes are created during the collision a secondary vertex, displaced from the primary one, can be observed. The impact parameters (d_0 and z_0) are defined as the distance, in the transverse and longitudinal plane respectively, between the nominal beam line and the track's point that is closest to it in the transverse plane. This point is called the perigee. The values of the transverse (d_0) and longitudinal (z_0) impact parameters (see figure 3.1) indicate to which vertex a track belongs. As such, the resolution on these measurements determines our ability to correctly associate tracks to primary and secondary vertices.

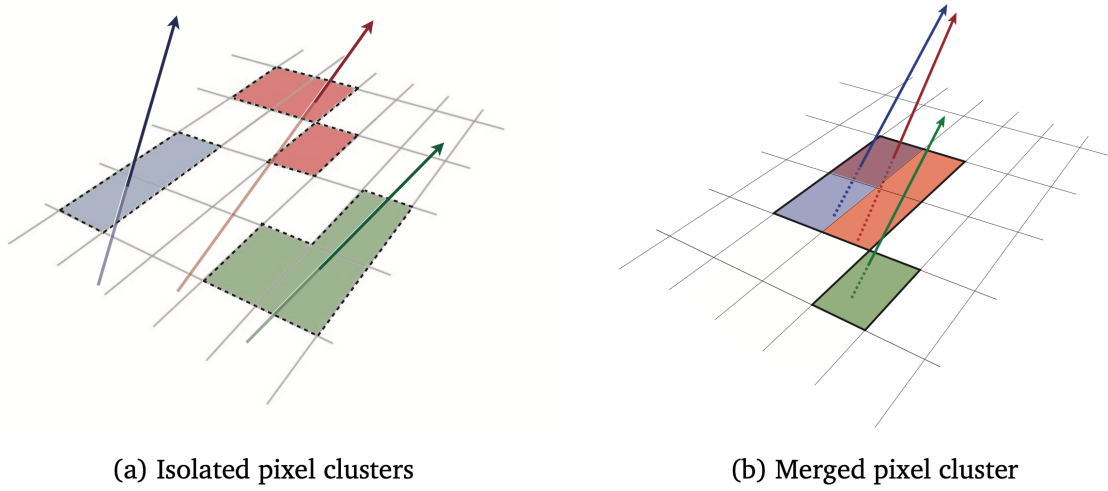


Figure 3.2: In (a), isolated pixel clusters from three resolved tracks. In (b), collimated tracks (dense/boosted environment) resulting in merged clusters [55].

3.2.2 Track reconstruction

The reconstruction of charged particles trajectories, known as tracks, is a challenging task and is performed in several steps, [55].

First, signals from neighbouring pixels or strips (from the SCT modules) are grouped together by using a connected component analysis (CCA). Their mean positions, which represent the point where the charged particle went through the detector, are used as three-dimensional space-points.

At high energies decay products from boosted massive particles are more and more collimated, which has an impact on track reconstruction. Indeed, if the charged particles are so close together that their separation distance is of the order of a few pixels, then the energy deposits in the tracker will start overlapping (as can be seen in figure 3.2). In this case the CCA algorithm will merge the hits and create a single *merged* cluster [56].

Iterative combinatorial track finding The three dimensional space-points, created in the clustering phase of the reconstruction, are used to seed the tracks. To ensure that many space-point combinations are considered (few points guarantee more combinations) but, at the same time, that it is possible to get a rough momentum estimation (many points are needed to give a good estimate), three space-points are used to create a seed. To guarantee that *purity* is high several requirements are placed at the different stages of track reconstruction. Purity is defined as the fraction of quality tracks over all tracks, where quality is a measure of how close to the truth the reconstructed track is. In particular, at the seeding stage, an additional space-point, compatible with the trajectory estimated from the seed, is required. Other requirements are put on the magnitude of the transverse momentum and the impact parameter of the seed track.

Purity varies depending on the nature of the space-points used to create the seed: seeds with only SCT space-points have the highest purity, followed by pixel only space-points. To optimise the efficiency of the track finding algorithm seeds are processed in order of purity and once a seed passes

all requirements, the space-points that compose it are removed from the iterative process.

A combinatorial Kalman filter builds track candidates from the seeds by incorporating the remaining space-points that are compatible with the preliminary trajectory, [55]. Multiple candidate tracks per seed are allowed, however, this results in a certain amount of ambiguity that needs solving.

Track candidates and ambiguity solving The track candidates are scored based on their quality and compared. If some ambiguity arises, the score is used to determine which track candidate is to be retained.

The quality of a track is established by evaluating several parameters. First of all, the more space-points that are assigned to the track, the higher the score, inversely, the higher the number of holes, the lower the score. The χ^2 of the track fit and the logarithm of its p_T value are also used to compute the score: energetic tracks with a good fit will have a higher score.

The ambiguity solver also has to deal with shared clusters, that is, clusters compatible with two or more track candidates. Such clusters are strong indicators of incorrect assignment and must therefore be limited. Clusters can be shared by at most two tracks, with preference for tracks with a high score. Furthermore a track cannot have more than two shared clusters: further clusters are removed from the candidate track, creating a stripped down version of it, which is then scored again.

Track candidates are rejected if their score is equal to zero, which automatically happens if they fail to satisfy basic quality criteria defined in terms of number of holes, shared clusters, large values of pseudorapidity, large values of impact parameter, etc. Finally, a high resolution fit is performed on all of the tracks remaining after all of the steps described above. The tracks that come out of the ambiguity solver are then extended to the TRT portion of the tracker⁽⁴⁾. A summary of the entire ambiguity solving process is shown in figure 3.3

3.2.3 Algorithmic Improvements to track reconstruction for boosted b-hadrons

The ambiguity solver described above is very robust, having been optimised in 2017 to deal with dense environments and the challenges they pose due to the highly collimated tracks. b-hadrons⁽⁵⁾ are long lived⁽⁶⁾ and when produced in an event (such as a W boson or top decay) they start decaying (and therefore interacting with the detector) at a short distance away from the interaction point. We say b-hadrons are displaced and refer to the point where they decay as secondary vertex (figure 3.4).

At high p_T the b-hadrons can be so displaced that the decay occurs after traversing a few layers of the inner detector. As seen in figure 3.5, tracks associated with the decay of a boosted b-hadron will typically have few hits in the innermost part of the detector and will have many shared hits in the SCT portion of the tracker due to the displacement of the secondary vertex. Both these factors

⁽⁴⁾We will not go into details about the TRT extension procedure since it is not relevant to the studies presented in this thesis, but a good description can be found in [57]

⁽⁵⁾Hadrons containing at least one b quark.

⁽⁶⁾Particles with lifetimes of the order of ps to ns.

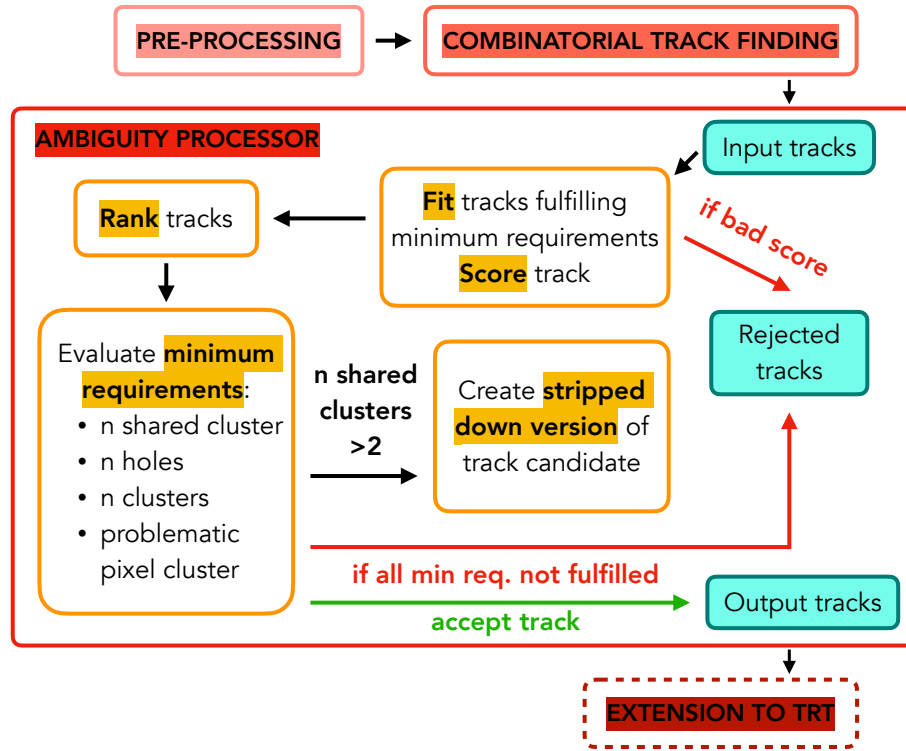


Figure 3.3: Summary of the track reconstruction procedure with a focus on the Ambiguity Solving step.

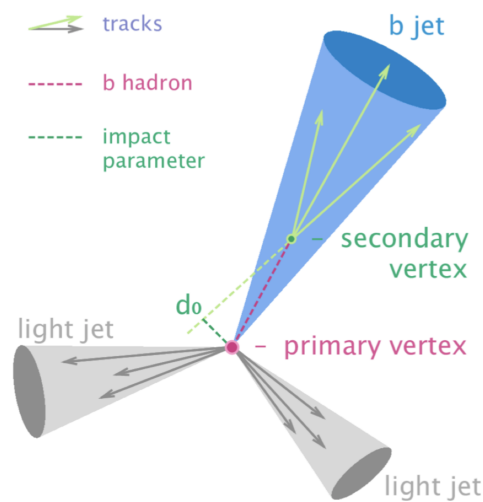


Figure 3.4: Anatomy of a decay resulting in a b-hadron and two light jets, such as a $t \rightarrow bqq$ decay [58]

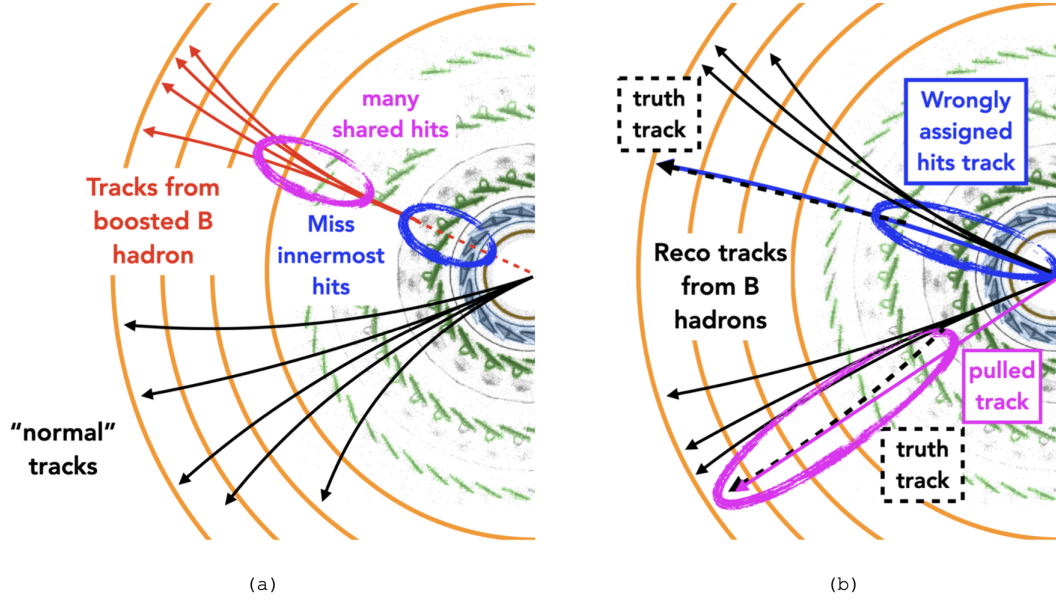


Figure 3.5: In (a) the difference between tracks coming from boosted b-hadrons and the “average” track. In (b) the effect of wrongly assigned hits on tracks coming from b-hadrons. The tracks are shown within the context of ATLAS tracking system: the green layers correspond to the pixel detector whereas the orange ones correspond to the SemiConductor Tracker (SCT) (see section 2.3).

mean that these tracks will either be rejected by the ambiguity solver or will carry wrongly assigned hits from the IBL and B layers of the tracker. This has a toll on our ability to correctly identify boosted B-hadrons.

The identification of b-hadrons is referred to as b-tagging. To achieve correct classification, several methods can be used. In particular the ATLAS collaboration uses a combination of many taggers⁽⁷⁾ to achieve the best performance. For flavour identification, taggers primarily rely on impact parameter measurement (used to determine the incompatibility with the primary vertex) and on secondary vertex reconstruction. The characteristics of tracks associated to a boosted b-hadron, described above, limit the resolution on the d_0 measurement and affect the reconstruction of the secondary vertex. The consequences of those track reconstruction inefficiencies for boosted b-tagging⁽⁸⁾ can be seen in figure 3.6⁽⁹⁾.

We can regain efficiency in track reconstruction and, at the same time, improve its quality. Indeed, most of the efficiency loss is due to the cuts applied in the ambiguity solver. At the same time, wrong hit assignments degrade the quality of the track. We therefore work in two directions: on one side loosening the cuts and on the other removing the innermost hits and seeing if the quality of the track

⁽⁷⁾Taggers are the algorithms used to identify the nature of the particle observed.

⁽⁸⁾The DL1r tagger is shown here, which is the primary b-tagging algorithm in ATLAS. It is a high-level tagger that takes as inputs the discriminants of simpler taggers.

⁽⁹⁾The background rejection also drops with increasing p_T [59].

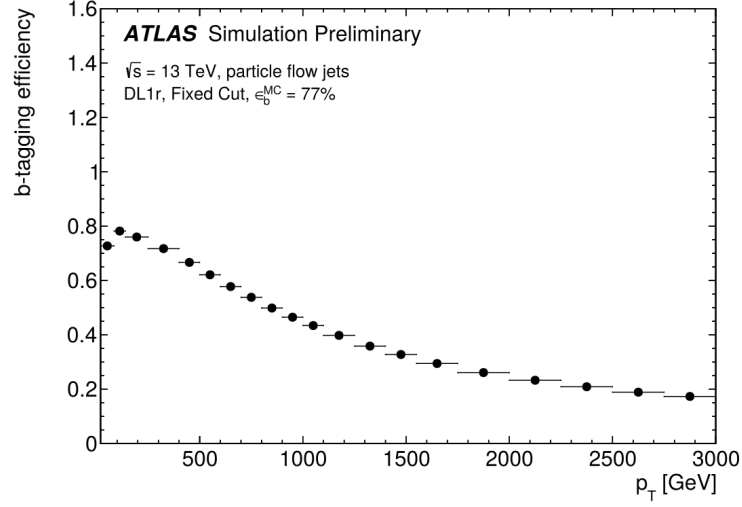


Figure 3.6: b-tagging efficiency using the DL1r tagger at the 77% working point. The efficiency drops at high p_T . [60]

improves.

We want to apply these changes only in a specific Region Of Interest (ROI). This is both to limit computing time, which is large for refitting the tracks, and to not increase the incidence of fake tracks. The aim of the ROI is to only apply the changes to b tracks, however we cannot identify such tracks with 100% certainty. Tracks belonging to the Region Of Interest have to meet certain criteria: first we identify topo-clusters in the jet with an energy larger than $E_{\text{cut}} = 150$ GeV, then we only select tracks that have a momentum larger than $p_T^{\text{cut}} = 15$ GeV and an angular distance between the track and the cluster smaller than $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.1$. These selections give us a good purity, meaning we mainly select tracks from b-hadrons, and a good efficiency, meaning we select most of the tracks that come out of b-hadrons. The validity of the ROI can be seen in figure 3.7.

Once the Region Of Interest has been defined the cuts on the number of shared hits, $N_{\text{shared}}^{\text{hits}}$, the number of non-shared hits, $N_{\text{non-shared}}^{\text{hits}}$, and the minimum number of silicon hits (that is both pixel and SCT hits), $N_{\text{Si}}^{\text{hits}}$, need to be optimised. The nominal values are changed as follows:

- The number of shared hits is allowed to be as high as 6 instead of the nominal value of 3.
- The minimum allowed number of non shared hits goes from 5 to 2.
- Tracks with as little as 6 silicon hits are accepted instead of the nominal 8.

The results obtained can be seen in figure 3.8. We observe a 15% increase in track efficiency for tracks coming from b-hadrons decaying after the Pix2 layer of the inner detector and a 10% gain for tracks coming from b-hadrons decaying after the B layer of the inner detector. Any increase in track efficiency should result in larger or equal gains in terms of b-tagging. Indeed, since each b-hadron generates roughly four charged particles (and therefore tracks), any gain in terms of track efficiency should result in a four fold gain in the tagger efficiency.

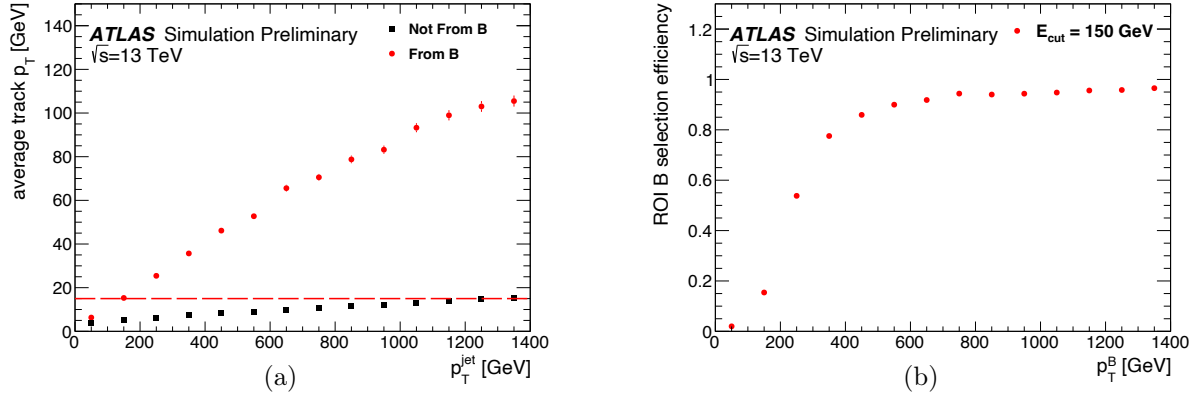


Figure 3.7: In (a), comparison of the average track p_T for tracks coming from b-hadrons and fragmentation tracks as a function of the jet p_T . In (b), comparison of the fraction of b-hadrons that are correctly contained by the ROI for different calorimeter energy deposit cut values, we want this number to be as close to one as possible. We consider a b as contained if all its decay products have a $p_T > 15$ GeV and are matched to a calorimeter energy deposit (with energy $> E_{cut}$) within $\Delta R < 0.1$ [61].

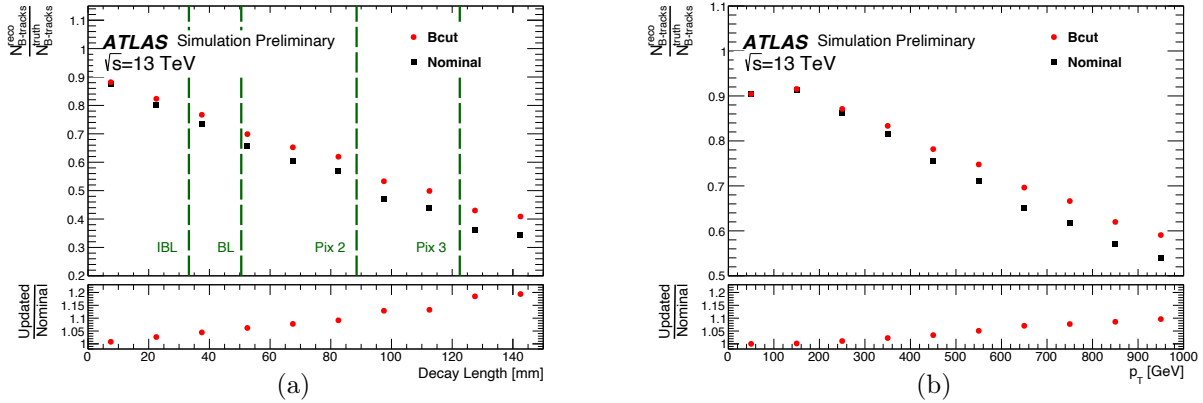


Figure 3.8: In (a) The per track efficiency as a function of decay length ($N_{B-tracks}^{reco}/N_{B-tracks}^{truth}$) of the Bcut sample (red) that requires a minimum number of non shared hits of 2 ($N_{non-shared}^{hits} \geq 2$), a maximum number of shared hits of 6 ($N_{shared}^{hits} < 7$) and minimum number of silicon hits to allow splitting of 7 ($N_{Si}^{hits} \geq 7$) is compared to that of the nominal sample (black), requiring $N_{non-shared}^{hits} > 6$, $N_{shared}^{hits} < 3$ and $N_{Si}^{hits} \geq 9$. The changes are applied only in a well defined Region Of Interest (ROI). The dark green lines indicate the location of the innermost layers of the ATLAS tracker. In (b) the same comparison of the per track efficiency as in (a) but as a function of p_T instead of decay length [61].

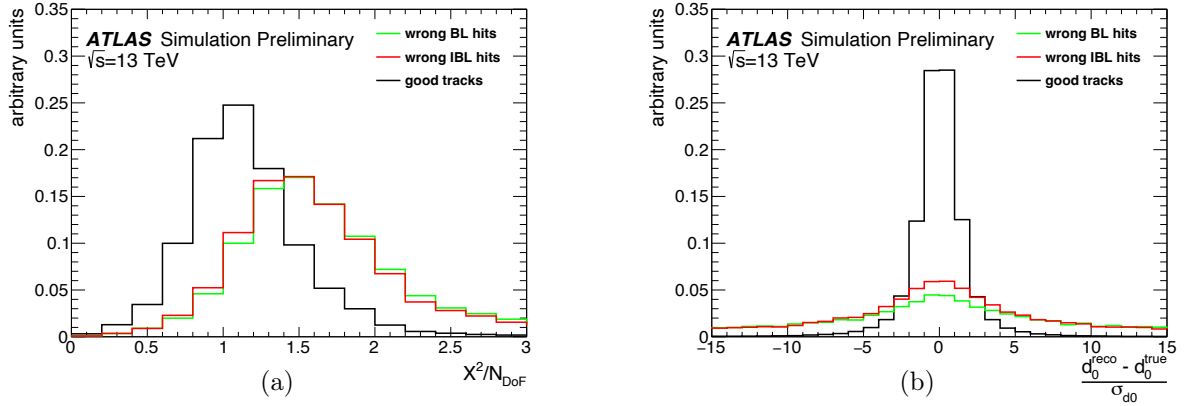


Figure 3.9: A comparison of properties for B tracks that have a wrongly assigned hit in the IBL and/or B-layer of the Inner Detector. In (a) such tracks (red and green) have a much worse quality of fit compared to tracks that don't have any wrongly assigned hit (good tracks). Such tracks also have a d_0 pull ($\frac{d_0^{\text{reco}} - d_0^{\text{true}}}{\sigma_{d_0}}$) distribution that is more spread out compared to tracks that don't have wrong hits (good tracks), as can be seen in (b) [61].

As can be seen in figure 3.9, the fact that a track carries hits that are wrongly assigned to it will result in two things. First, the “goodness” of the fit for those tracks decreases: for well reconstructed tracks the value of $\chi^2/NDOF$ should be close to 1 whereas the same value for tracks with wrongly assigned hits tends to be larger. Secondly, the tracks’ pull, that is the difference between the reconstructed and true impact parameter, divided by the uncertainty on the measurement, worsens. For well reconstructed tracks, the distribution of this value should be a gaussian centred around 0 and with a standard deviation of 1. Tracks with incorrect hits, however, will have a reconstructed impact parameter that is far from the true value and, therefore, in a pull value much larger or smaller than zero. This effect will result in a wider pull distribution that does not correspond to that of well reconstructed tracks. The idea is that the wrongly assigned hits in the innermost layers of the detector dramatically change the reconstructed secondary vertex (as shown in figure 3.5) and, as a consequence, the reconstructed value of the impact parameter.

To improve the impact parameter measurement, we target the tracks in the Region Of Interest, which we recursively strip down by removing their innermost hits. We fit these stripped down versions and look at the χ^2 measure described above. If by stripping down the track we achieve a better fit, we keep the stripped down version, otherwise, the original one. We only allow the removal of hits up until the Pix2 layer. As seen in figure 3.10, with this method we can reduce by up to 50% the number of wrongly assigned IBL hits, up to 25% the number of wrongly assigned B hits, and an overall 10% when looking at all the pixel hits. The benefits of employing both of the algorithmic improvement strategies, that is loosening the cuts and refitting the track, are shown with a red marker, the result of using one or the other are depicted in blue and green respectively, finally the black markers show the nominal track reconstruction.

More studies need to be made in order to truly evaluate the impact of these two courses of action.

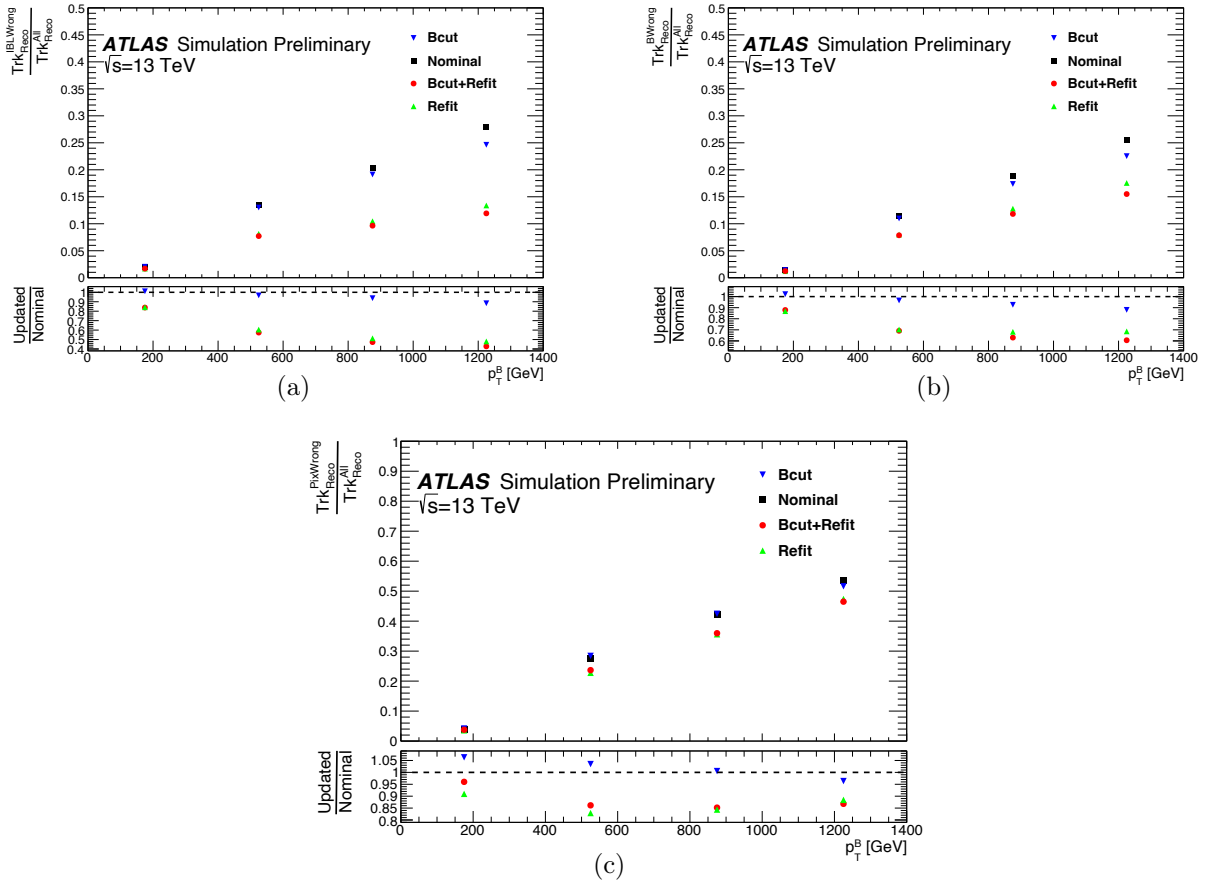


Figure 3.10: The ratio, per B, of (a) the number of reconstructed tracks with a wrongly assigned IBL hit and the total number of reconstructed tracks for 4 different reconstruction strategies, (b) the same ratio for tracks with a wrongly assigned B-layer hit, and in (c) the ratio for tracks with at least one wrongly assigned pixel hit [61].

In particular these reconstruction strategies should be tested in a full reoptimisation of the b-taggers.

3.3 Particle Flow

We have now described the objects reconstructed by the calorimeter and the tracker portions of the detector. We now introduce the concept of Particle Flow. The idea is to exploit the measurements of all the subdetectors composing an experiment in order to reconstruct individually each particle produced in the collision. This approach, when successful, leads to better performance in jet reconstruction and Missing Transverse Momentum (MET) and also improves the identification of charged leptons (electrons, muons and taus). For a particle flow algorithm to be efficient, the detector must be able to correctly separate charged and neutral energy contributions in hadronic showers. To achieve this, a large magnetic field, a calorimeter with high granularity and efficient tracking are of primary importance.

The ATLAS detector wasn't designed specifically for particle flow, instead each subdetector was optimised to obtain a certain resolution on key measurements. However, with the search for new particles moving towards unprecedented energies and increasing pile-up conditions, this idea of independent objects, each coming out of different experiment parts is no longer sufficient.

Taking the example of the $VV \rightarrow JJ$ analysis, the two quarks produced in the decay of a massive vector boson ($V \rightarrow qq$) become increasingly collimated as the parent particle becomes more energetic ($\Delta R_{qq} \approx 2m_V/p_T^V$). Hence, we reach a regime, the highly boosted regime ($p_T^V > 300$ GeV), where the granularity of the ATLAS calorimeter isn't sufficient to resolve the angular separation between the two quarks. The spatial resolution of the calorimeter therefore becomes a limiting factor in our ability to distinguish between the hadronic decay of a vector boson and the overwhelming QCD background. Indeed, new physics searches in highly boosted regimes rely heavily on the angular and energy resolution of the inputs to the jet reconstruction algorithm (see section 4.1).

At high energies, even though its angular resolution is insufficient, the ATLAS calorimeter has excellent energy resolution (section 2.4). On the other hand, the tracker angular resolution is excellent, especially at high p_T where, since the tracks are less bent by the magnetic field, the extrapolation to the calorimeter becomes increasingly precise. At the same time, since the p_T of the tracks is obtained from the curvature, at such energies, the tracker's p_T resolution decreases (section 2.3). Therefore, using information from both the tracker and the calorimeter allows us to build a combined object with balanced angular and energy resolution to be used as a building block for jet reconstruction: the Track-CaloCluster (see section 3.3.1).

Particle flow can also be used to build inputs to jet reconstruction that allow us to improve the performance and pile-up stability in low p_T jets. Particle Flow Objects (PFOs), described in section 3.3.2, exploit the superior energy resolution of the tracker in the low energy regime. The tracker information describes the properties of charged particles with great precision while the energy left in the calorimeter after the energy subtraction algorithm, is used to reconstruct neutral particles.

Finally these two objects can be combined in a third object, called United Flow Object (UFO) (see section 3.3.3) that takes advantage of the increased performance at low p_T of PFlow objects and, at the same time, the excellent behaviour of TCCs at high p_T , to provide a single optimal definition

everywhere.

3.3.1 Track-CaloClusters

For Track-CaloClusters (TCCs) the idea is to work at the level of individual tracks and clusters to define new objects to use as input for jet reconstruction. By exploiting information from both the tracker and the calorimeter the aim is to further improve the resolution on jet measurements.

The concept behind TCCs is simple, however their mathematical implementation can become quite complex when applied to real events. It is therefore useful to go through a few examples with increasing complexity before giving a rigorous mathematical definition.

To be used as inputs in jet reconstruction algorithms, the TCCs must be defined as four-vectors: $\mathbf{TCC} = (p_{TCC}, \eta_{TCC}, \phi_{TCC}, m_{TCC})$. In the following equations the numbering of scenarios, clusters and tracks follows the one in figure 3.11, for simplicity of notation in the equations, clusters will be represented by the letter c and tracks will be represented by t both followed by their label number. Four-vectors are indicated by a bold font.

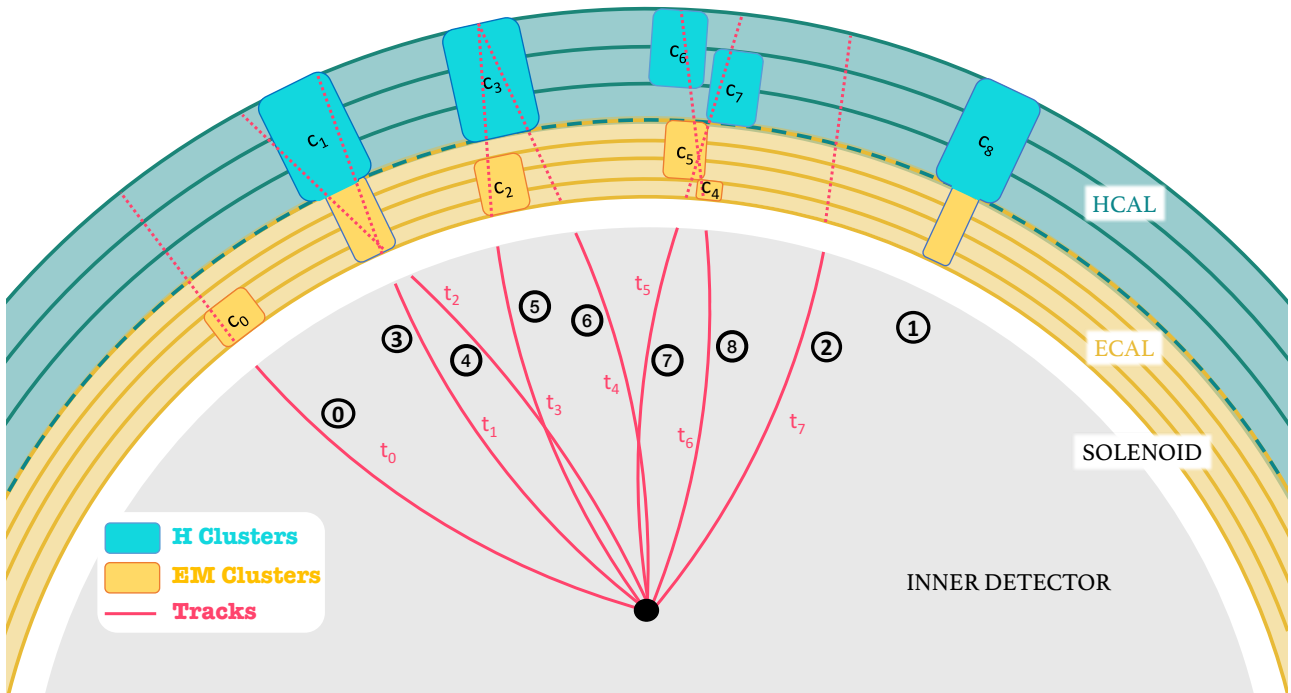


Figure 3.11: Different TCC scenarios. The circled numbers correspond to the TCC objects, the tracks and clusters are also labeled in order to make the equations clearer.

Examples The simplest case for a TCC corresponds to scenario ①. We have one track that is matched to one cluster, as mentioned before we want to get a better overall resolution by exploiting the superior angular resolution of the tracker and combining it with the excellent energy resolution of

the calorimeter: the Lorentz vector is built by taking the η and ϕ angles of the track and the p_T and m of the cluster, the latter is null because clusters have their mass set to 0.

$$\mathbf{TCC}_0 = (p_T^{c0}, \eta^{t0}, \phi^{t0}, m^{c0} = 0) \quad (3.3)$$

Single cluster In this case, the cluster isn't matched to any track, as in scenario (1). We therefore simply take the cluster's four-vector:

$$\mathbf{TCC}_1 = (p_T^{c8}, \eta^{c8}, \phi^{c8}, m^{c8} = 0) \quad (3.4)$$

Single track The tracks can also be unmatched, as in scenario (2), and as for the single cluster we simply take the track's four-vector:

$$\mathbf{TCC}_2 = (p_T^{t7}, \eta^{t7}, \phi^{t7}, m^{t7}) \quad (3.5)$$

The simple cases have been treated, however scenarios can get complex quite quickly. In the next scenarios we have energy sharing: clusters have contributions from multiple tracks which can, at the same time, contribute to the energy of several clusters. To account for this energy sharing we apply weights to the energy (p_T) of the clusters. We assume that the split contributions in energy are proportional to the tracks' and clusters' p_T .

- TCC (3) and (4) : The cluster has contributions from both tracks, the split is proportional to the p_T of each track.

$$\mathbf{TCC}_3 = \left(\frac{p_T^{c1} \cdot p_T^{t1}}{p_T \text{ of } (\mathbf{t}_1 + \mathbf{t}_2)}, \eta^{t1}, \phi^{t1}, m = 0 \right) \quad (3.6)$$

$$\mathbf{TCC}_4 = \left(\frac{p_T^{c1} \cdot p_T^{t2}}{p_T \text{ of } (\mathbf{t}_1 + \mathbf{t}_2)}, \eta^{t2}, \phi^{t2}, m = 0 \right) \quad (3.7)$$

- TCC (5) and (6) : The amount of energy going into cluster 3 depends on how much energy from track 3 went into cluster 2. We therefore have:

$$\mathbf{TCC}_5 = (p_T^{P5}, \eta^{t3}, \phi^{t3}, m^{P5} = 0) \quad (3.8)$$

and

$$\mathbf{TCC}_6 = (w_{t4}^{c3} \cdot p_T^{c3}, \eta^{t4}, \phi^{t4}, w_{t4}^{c3} \cdot m^{c3} = 0) \quad (3.9)$$

Where \mathbf{P}_5 is the four-vector, associated to \mathbf{TCC}_5 , we get by combining the weighted four-vectors of the two clusters matched to track 3, and w_{t4}^{c3} is a weight: for each track-cluster matched pair a weight is computed to account for energy sharing with other tracks and clusters, we therefore

label the weights by indicating which cluster c and track t it was computed for.

$$\mathbf{P}_5 = w_{t_3}^{c_3} \cdot \mathbf{c}_3 + w_{t_3}^{c_2} \cdot \mathbf{c}_2$$

and

$$w_{t_3}^{c_3} = \frac{p_T^{t_3}}{p_T \text{ of } (\frac{p_T^{c_3}}{p_T \text{ of } (\mathbf{c}_3 + \mathbf{c}_2)} \cdot \mathbf{t}_3 + \mathbf{t}_4)} \cdot \frac{p_T^{c_3}}{p_T \text{ of } (\mathbf{c}_3 + \mathbf{c}_2)},$$

$$w_{t_3}^{c_2} = 1, \quad w_{t_4}^{c_3} = \frac{p_T^{t_4}}{p_T \text{ of } (\frac{p_T^{c_3}}{p_T \text{ of } (\mathbf{c}_3 + \mathbf{c}_2)} \cdot \mathbf{t}_3 + \mathbf{t}_4)}$$

- TCC (7) and (8) : This case is the one closest to reality, we follow the same procedure we used for \mathbf{TCC}_5 and \mathbf{TCC}_6 . We have:

$$\mathbf{TCC}_7 = (p_T^{P_7}, \eta^{t_5}, \phi^{t_5}, m^{P_7} = 0) \quad (3.10)$$

and

$$\mathbf{TCC}_8 = (p_T^{P_8}, \eta^{t_6}, \phi^{t_6}, m^{P_8} = 0) \quad (3.11)$$

Where \mathbf{P}_7 and \mathbf{P}_8 are the four-vector, associated to \mathbf{TCC}_7 and \mathbf{TCC}_8 , we get by combining the weighted four-vectors of the clusters matched to track 5 and 6 respectively.

$$\mathbf{P}_6 = w_{t_5}^{c_5} \cdot \mathbf{c}_5 + w_{t_5}^{c_7} \cdot \mathbf{c}_7,$$

$$\mathbf{P}_7 = w_{t_6}^{c_4} \cdot \mathbf{c}_4 + w_{t_6}^{c_5} \cdot \mathbf{c}_5 + w_{t_6}^{c_6} \cdot \mathbf{c}_6$$

The weights become more intricate with the number of clusters and tracks matched to each other (cf. TCC_5 compared to TCC_3 or TCC_4), therefore, without writing explicitly the mathematical expressions of P_7 and P_8 , one can see how complexity quickly grows when it comes to more realistic topologies.

No matter how complex the scenarios, there truly are only three different classes, or *tastes*. These tastes are assigned depending on whether or not the matching between a track and a cluster was successful.

- The first taste is *neutral*, corresponding to the case where a calorimeter cluster isn't matched to any track, TCC (1).
- The second taste is *charged*, corresponding to the case where a track isn't matched to any cluster, TCC (2).
- The third taste is *combined*, corresponding to the cases where there is at least one match between the seed track and a cluster. This is the taste of all the other scenarios studied in the previous section.

As shown in the examples we have built an algorithm to handle energy sharing:

- We want to build a TCC for each track in the event. The track for which the TCC is built is defined as the seed track.

- We match clusters and tracks to one another.
- Now we need to build the actual TCC four-vector. As said before, we use the angular coordinates from the track and the transverse momentum and mass from a combined object. The combined object is the sum of the four-vectors of the clusters matched to the track scaled by a weight that accounts for energy sharing.
- For each cluster matched to the seed track, the energy is split between the tracks that match it. The split is defined as the p_T fraction that each track contributes to the cluster compared to that of the other matching tracks.

Therefore, a TCC is associated to every seed track τ in the event, the seed track is matched to any number of clusters C_τ which are, in turn, matched to any number of tracks T_c that are, once again, matched to any number of clusters C_t . The algorithm handles the energy sharing between all these matched objects and can be summarised in an equation. We indicate by X_y the group of matched X objects to a given y object, where the X matched objects are C clusters if the given y object is a track, and, vice versa, T tracks if y is a cluster.

The energy sharing procedure is based on the assumption that the split in energy is proportional to the transverse momentum of the objects involved and therefore makes use of three general concepts. First, the clusters, C_τ , matched to the seed track τ , each contribute to the resulting TCC by an amount that is proportional to their p_T fraction, f_τ^c , out of all matched clusters. Second, each cluster c contributes to all the TCC objects seeded by the tracks, T_c , it is matched to. Its contribution to TCC_τ should therefore be weighted by the fraction of p_T demanded by the seed track τ compared to all other tracks t matched to the cluster, $F_{c,t}^\tau$. Finally, the contribution of each of those track t , should be proportional to the fraction of energy the cluster c represents compared to all other cluster matching the track, f_t^c , [62].

$$f_\tau^c = \frac{p_T^c}{p_T \left[\sum_{k \in C_{\tau}} \mathbf{p}^k \right]}, \quad F_{c,t}^\tau = \frac{p_T^\tau}{p_T \left[\sum_{t \in T_c} \mathbf{p}^t f_t^c \right]}, \quad f_t^c = \frac{p_T^c}{p_T \left[\sum_{k \in C_t} \mathbf{p}^k \right]} \quad (3.12)$$

In summary, f_τ^c is an indication of how relevant c is to TCC_τ compared to all the topo-clusters matching the seed track τ . If the cluster has a small transverse momentum compared to the others it will have a small weight and will thus have a small impact on the final result. Similarly, $f_{c,t}^\tau$ shows the relevance of the seed track τ compared to all the tracks matching the cluster c . This is essentially a normalisation factor that allows us to determine the amount of energy each cluster is able to contribute to each TCC seeded by a track matching it. The final TCC energy sharing equation is:

$$\mathbf{M}_\tau = \sum_{c \in C_\tau} \mathbf{p}^c f_\tau^c F_{c,t}^\tau = \sum_{c \in C_\tau} \mathbf{p}^c \frac{p_T^c}{p_T \left[\sum_{k \in C_{\tau}} \mathbf{p}^k \right]} \frac{p_T^\tau}{p_T \left[\sum_{t \in T_c} \mathbf{p}^t \frac{p_T^c}{p_T \left[\sum_{k \in C_t} \mathbf{p}^k \right]} \right]} \quad (3.13)$$

\mathbf{M}_τ is a four-vector that accounts for energy sharing and with which we define the transverse momentum and the mass of the TCC. The angular components come from the seed track τ .

$$\mathbf{TCC}_\tau = (p_T \text{ of } \mathbf{M}_\tau, \eta^\tau, \phi^\tau, m \text{ of } \mathbf{M}_\tau) \quad (3.14)$$

It is interesting to point out that the transverse momentum of the tracks only enters the equations through ratios. As such, the TCC procedure isn't affected to first order by the poor transverse momentum resolution of the tracker at high p_T since the limiting factors cancel each other out.

3.3.2 PFlow

The ATLAS experiment has developed another particle flow algorithm to be used in jet reconstruction. The use of this approach in ATLAS is especially useful to improve performance and pile-up stability in low p_T jets. The high luminosity reached by the LHC in Run 2 introduces a further complication in event reconstruction: in this environment, multiple interactions can contribute to the detector signals associated with a single bunch-crossing (pile-up). These interactions, which are mostly soft, have to be separated from the hard interaction that is of interest, [63]. To reduce pile-up the tracker is crucial as the tracking procedure assigns the tracks to a vertex which can be the primary (hard scatter) one, or a pile-up one⁽¹⁰⁾. Therefore, using information from both the tracker and the calorimeter to create a combined object makes it possible to identify charged pile-up in the calorimeter (clusters only associated to pile-up tracks). At the same time, the acceptance of the detector is extended to softer particles coming from hard-scatter vertices⁽¹¹⁾, this allows reconstruction of soft activity, which is important for the calculation of missing transverse momentum in the event, [64]. However, as mentioned earlier, PFlow in ATLAS is mainly useful in the low p_T region. In fact, at low p_T the tracker p_T resolution is better than the calorimeter energy resolution, as is the angular resolution for a single reconstructed particle.

In practice Particle Flow Objects are created using a cell-based energy subtraction algorithm: the track's energy (that is the energy deposited by the particle that produced the track) is subtracted cell by cell from the set of matched topo-clusters, [63]. The remaining energy in the calorimeter is then used to reconstruct neutral particles or removed if it is within the expected shower fluctuations of a single particle's signal. The entire procedure is illustrated in 3.12.

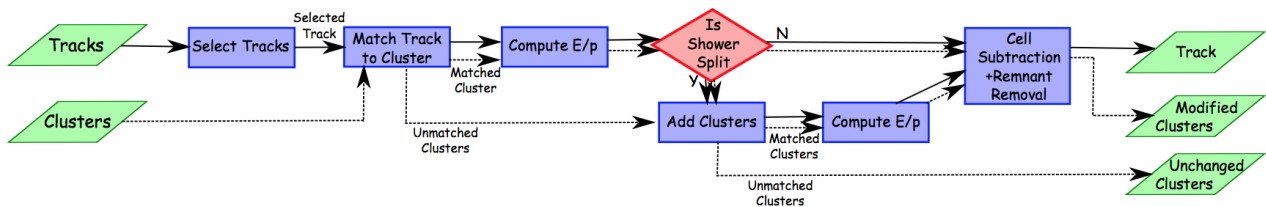


Figure 3.12: A flow chart of how the particle flow algorithm proceeds [63].

⁽¹⁰⁾A primary vertex is any vertex consistent with the beam. However, in this case, we label a vertex as primary if it is the vertex with the largest value for the sum $\sum p_T^2$, where the sum is over all the tracks associated with the vertex.

⁽¹¹⁾tracks are reconstructed for objects with a minimum transverse momentum of 500 MeV, objects that soft often don't pass the noise threshold applied in the calorimeter 2.2

Since the track's energy is subtracted from the calorimeter, this methodology requires excellent p_T resolution from the tracker, which we only have at low p_T . To account for this effect, the energy subtraction procedure is gradually turned off with increasing p_T . Since it attempts to match the track to a single topo-cluster the procedure also relies on the topological clustering algorithm and its ability to contain a single particle's shower in a single topo-cluster. Since this is not always the case, and the shower is often split over two topo-clusters, the particle flow algorithm needs to be able to handle these scenarios. This is done by first determining if, indeed, the shower is split across several clusters and then by merging the clusters over which the shower is split. To identify such cases, a threshold on the significance of the difference between the expected energy and that of the matched topo-cluster is used, the split shower recovery procedure is applied if:

$$S(E^{clus}) = \frac{E^{clus} - \langle E_{dep} \rangle}{\sigma E_{dep}} < -1 \quad (3.15)$$

3.3.3 United Flow Objects

Finally, United Flow Objects (UFOs), [65] are a combination of PFOs and TCCs. They are built with the aim of benefiting from the complementary behaviour of these two objects in different p_T ranges, thus creating a new superior object.

The various steps of the reconstruction algorithm can be seen in figure 3.13. The algorithm begins by creating all the Particle Flow Objects in the event and then proceeds to remove the ones associated to pile-up tracks. The remaining PFOs are split into charged PFOs with energy subtraction applied (PFO_{sub}), charged PFOs without energy subtraction (PFO_{nosub}) due to their high momentum or to the fact that they are located in a dense environment and, finally, neutral PFOs ($PFO_{neutral}$). A modified TCC splitting is applied to the latter two type of PFOs: only tracks from the hard-scatter vertex are used in order to maintain pile-up robustness. Furthermore, tracks already matched to a PFO_{sub} are removed from the TCC splitting procedure as they have already been well assigned to topo-clusters and their energy contribution in the calorimeter has already been subtracted. The TCC algorithm then proceeds as usual on the reduced collection of tracks and PFOs (neutral and unsubtracted).

This methodology produces the best results, maintaining both the PFO subtraction at low p_T and the benefits obtained with TCC splitting at high p_T (see sections 5.3 and 7.5).

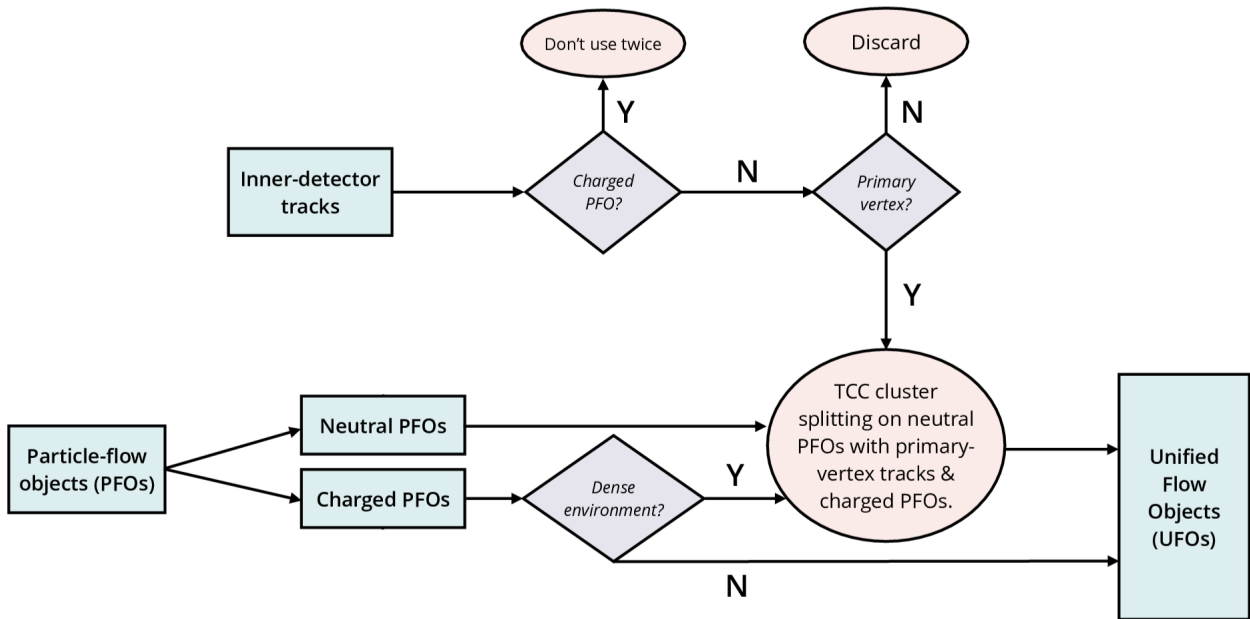


Figure 3.13: A flow chart of how the United Flow Object algorithm proceeds [65].

4

Jet reconstruction

Contents

| | | |
|------------|--------------------------------------|-----------|
| 4.1 | Jet reconstruction algorithm | 59 |
| 4.2 | Inputs to jet reconstruction | 62 |
| 4.3 | Jet grooming | 63 |
| 4.3.1 | Trimming | 63 |
| 4.3.2 | SoftDrop | 64 |
| 4.4 | Pile-up mitigation techniques | 65 |
| 4.4.1 | Constituent subtraction | 65 |
| 4.4.2 | SoftKiller | 66 |
| 4.5 | Jet calibration | 66 |
| 4.5.1 | Simulation based calibration | 66 |
| 4.5.2 | <i>In situ</i> energy calibration | 67 |
| 4.5.3 | <i>In situ</i> jet mass calibration | 68 |
| 4.6 | Jet substructure variables | 68 |
| 4.6.1 | Mass | 69 |
| 4.6.2 | Energy correlation variables | 70 |
| 4.6.3 | N-subjettiness | 71 |
| 4.6.4 | Other substructure variables | 71 |

One of the characteristic features of the LHC is the production of highly collimated spray of hadrons, called **jets**, that originate from quarks and gluons produced in the proton proton collisions. Indeed, quarks and gluons cannot exist individually: as described in section 1.1.2 the coupling g_s of the strong force increases with distance. This causes the fragmentation and hadronisation of the primary particles resulting in colourless, due to colour confinement, bound states: hadrons.

Jets are as close as we can get to a physical single hard quark or gluon.

The jets studied in this thesis originate from the hadronic decay of vector bosons, see section 1.4.1. Each vector boson would usually be reconstructed as two jets, one for each quark. However, because of the high energy reached at the LHC, W and Z bosons can be produced with momentum, p_T^V , much larger than their mass, m_V . The distance between the decay products of such boosted bosons becomes

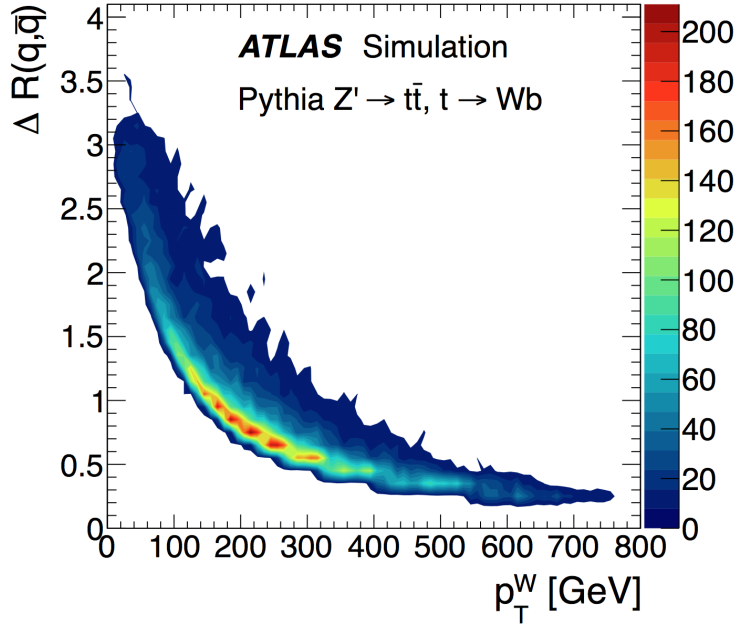


Figure 4.1: Angular distance between the light quark and anti-quark from W decays in $t \rightarrow Wb$ events as a function of the p_T of the W boson (p_T^W). The distribution is at the generator level and does not include effects due to initial and final-state radiation, or the underlying event [66].

smaller with the increase of the vector boson p_T . Such dependency is expressed through an empiric formula:

$$\Delta R_{qq} = \sqrt{\Delta\phi_{qq}^2 + \Delta\eta_{qq}^2} \approx \frac{1}{\sqrt{x(1-x)}} \frac{m_V}{p_T^V} \quad (4.1)$$

where x and $(1-x)$ represent the fraction of momentum carried by each quark respectively. In W, Z decays the fraction of momentum carried by each quark is approximately the same, and the formula reduces to $\Delta R_{qq} \approx \frac{2m_V}{p_T^V}$.

Figure 4.1 clearly shows that, already at 300 GeV, the majority of the W bosons decay products have an angular separation of around 0.5. Hence, the ability to resolve the two quarks by using standard narrow jets, $R = 0.4$, quickly degrades: the showers originating from the two quarks start to overlap and cannot be fully resolved. Large- R jets with $R = 1.0$, are then used to make sure that all the energy deposited from the hadronic decays is contained within the jet. We can therefore separate hadronic boson decays in two regimes, resolved and boosted: for $p_T < 200$ GeV the vector boson decays are reconstructed into two resolved narrow jets, when the transverse momentum of the parent particle is larger than 200 GeV a single large- R ($R = 1.0$) jet is built to contain both decay products.

In the following we will refer to samples containing jets originating from light quarks or gluons as QCD, di-jet or background samples, whereas samples containing jets originating from vector bosons will be denoted as signal or W' samples.

The high luminosity conditions of the LHC create an environment where multiple interactions can contribute to the detector signals associated with a single bunch-crossing. This phenomenon is called

pile-up and has to be accounted for when building jets. Pile-up interactions, which are mostly soft, have to be separated from the hard scatter interaction of interest. Pile-up contributions can be divided in two categories: *in-time pile-up* where the contributions come from proton-proton collisions of the same bunch-crossing, and *out-of-time pile-up* that consists of energy remnants in the detector of previous bunch-crossings, [63]. High p_T large- R jets originating from light quarks or gluons have significantly different characteristics than jets originating from vector bosons. Hence to gain insight on the nature of large- R jets substructure techniques can be used. Generally, W or Z associated jets will present two high energy regions, corresponding to the two quarks, of approximately same p_T with a large ΔR between them resulting in a large mass. The p_T balance between the two prongs and the angular distance between them will however differ between a longitudinally polarised vector boson and a transversally polarised one (as will be further detailed in section 7.1). On the other hand, QCD jets are characterised by a single dense core of energy surrounded by softer elements. The removal of pile-up and underlying event through *grooming* techniques and *pile-up mitigation* techniques allows one to better resolve such substructure properties, see section 4.3.

Substructure properties are defined and described in the last part of this chapter.

4.1 Jet reconstruction algorithm

Jet reconstruction algorithms aim to correctly reconstruct the sprays of hadrons that originate from quarks or gluons by grouping together particles according to a set of rules [67]. The radius parameter of the jet algorithm, R , determines if two inputs should be associated to the same jets. Furthermore, a recombination scheme, that is, the procedure that determines how particles should be combined, must be defined. The most common procedure, and the one used in the algorithms hereby described, consists of adding the four-vector of the individual particles together.

Summarising, jets are built by combining four-vectors (inputs) following a given set of rules (algorithm) within a specified radius R . Any of these three parameters can be changed and will result in a different collection of jets.

For this reason, a general set of recommendation should be followed when introducing a new jet definition, as decided by the "Snowmass accord" in 1990, [68]:

- It needs to be simple to implement for experimental analysis
- It needs to be simple to implement for theoretical calculations
- It needs to be defined at any order of perturbation theory
- It has to yield finite cross sections at any order of perturbation theory
- It has to yield a cross section that is relatively insensitive to hadronisation

The last three points are now comprised in the concepts of infrared and collinear safety of the algorithm. An infrared and collinear safe algorithm guarantees that the addition of a soft emission (low p_T gluon), as shown in 4.2 or the introduction of a collinear (small angle) splitting, as shown in figure 4.3, does not change the set of hard jets found in the event.

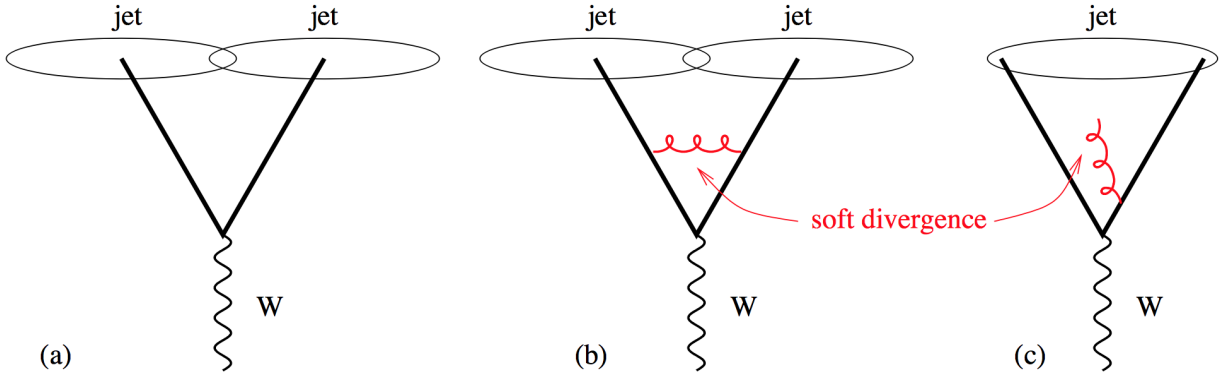


Figure 4.2: Configurations illustrating IR unsafety in events with a W and two hard partons. The addition of a soft gluon converts the event from having two jets to just one jet [67].

This is especially important due to the fact that both collinear splitting and infrared emission occur randomly throughout perturbative processes and, therefore, cannot be predicted. An example of a collinear unsafe algorithm is the most intuitive algorithm one could think of: define the seed particles of jet reconstruction as the hardest particle and group together all the particles within a fixed radius from such a seed particle. If the seed particle undergoes a splitting, its momentum might not be as high and the seed may change, and with it the entire jet.

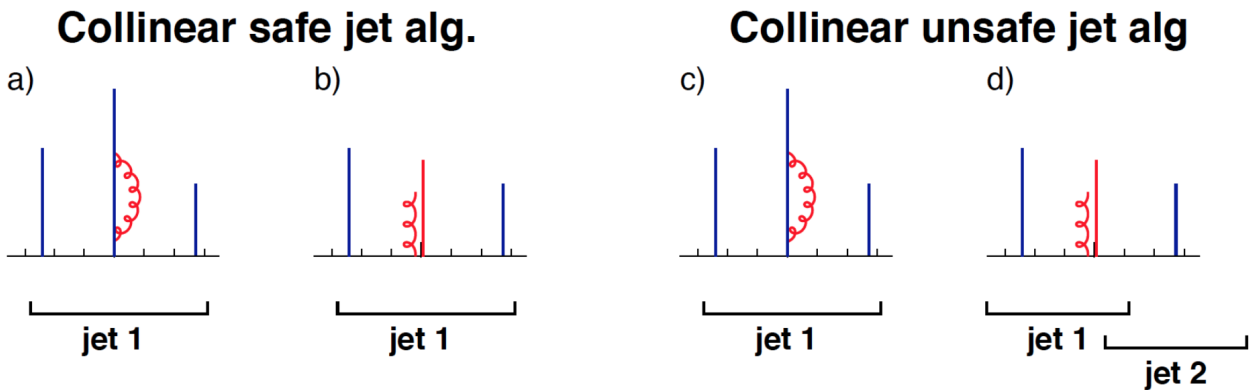


Figure 4.3: Illustration of collinear safety (left) and collinear unsafety (right). Partons are vertical lines, their height is proportional to their transverse momentum, and the horizontal axis indicates rapidity [67].

In this section we will describe three of the main methodologies used: the Cambridge-Aachen (C/A) [69], Anti- k_t [70] and k_t [71] algorithms. All three are collinear and infrared safe and are sequential reconstruction algorithms⁽¹⁾. The idea behind such algorithms is to iterate over all inputs (that are

⁽¹⁾As opposed to cone algorithms, one of which was described in the example above. One of their main feature is to have a rigid circular boundary which means that their shape is insensitive to additional soft particles.

the four-vectors of any object desired) and to combine them based on some distance parameter. Two distance parameters must be defined: d_{ij} , the distance between the two particles i and j , and d_{iB} , the distance between entity i and the beam axis (B).

$$d_{ij} = \min(p_{T_i}^{2p}, p_{T_j}^{2p}) \frac{\Delta R_{ij}^2}{R_{jet}^2}$$

$$d_{iB} = p_{T_i}^{2p}$$

where R_{jet} is the radius parameter chosen in the jet definition, and

$$\Delta R_{ij} = \sqrt{(\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2} \quad (4.2)$$

is the angular separation between the two i, j particles. If the distance between the two entities i, j , is smaller than the distance between particle i and the beam axis, i and j are combined and treated as a single entity and, as such, added to the list of inputs on which the algorithm will reiterate. If, on the other hand, there is no particle j closer to i than the beam, i is qualified as a jet and removed from the list of inputs. The algorithm reiterates until there are no more entities in the inputs list. The free parameter of the algorithm p is what defines the recombination order and, therefore the reconstruction algorithm: $p = 0$ for Cambridge-Aachen (C/A), $p = 1$ for k_t and $p = -1$ for Anti- k_t .

The C/A algorithm doesn't take into account the energy of the particles it recombines, only the angular distance between particles is considered. On the other hand k_t favours the reclustering of softer particles first and then recombines such soft entities to harder particles later in the iteration process. Both algorithms described therefore provide, through their clustering history, spatial and kinematic information about the substructure of the jets [66]. However, because of the recombination order, both algorithms result in jets whose shapes aren't stable and vary when soft particles are added. Anti- k_t , on the other hand, clusters soft particles to hard ones long before they cluster among themselves; this results in stable, circular jets. In this case however, the substructure classification that derives from the clustering-sequence carries little information about the p_T ordering of the shower (as k_t does) or the wide-angle structure (as C/A does), [67].

The jets resulting from the different jet algorithms are shown in figure 4.4.

To exploit both the stability against pile-up of the Anti- k_t algorithm and the substructure information resulting from the k_t algorithm, the former is used to build the jets, then, the latter is used to recluster the jet's constituents to enable the calculation of other properties such as the k_t ordered splitting scales, 4.6.

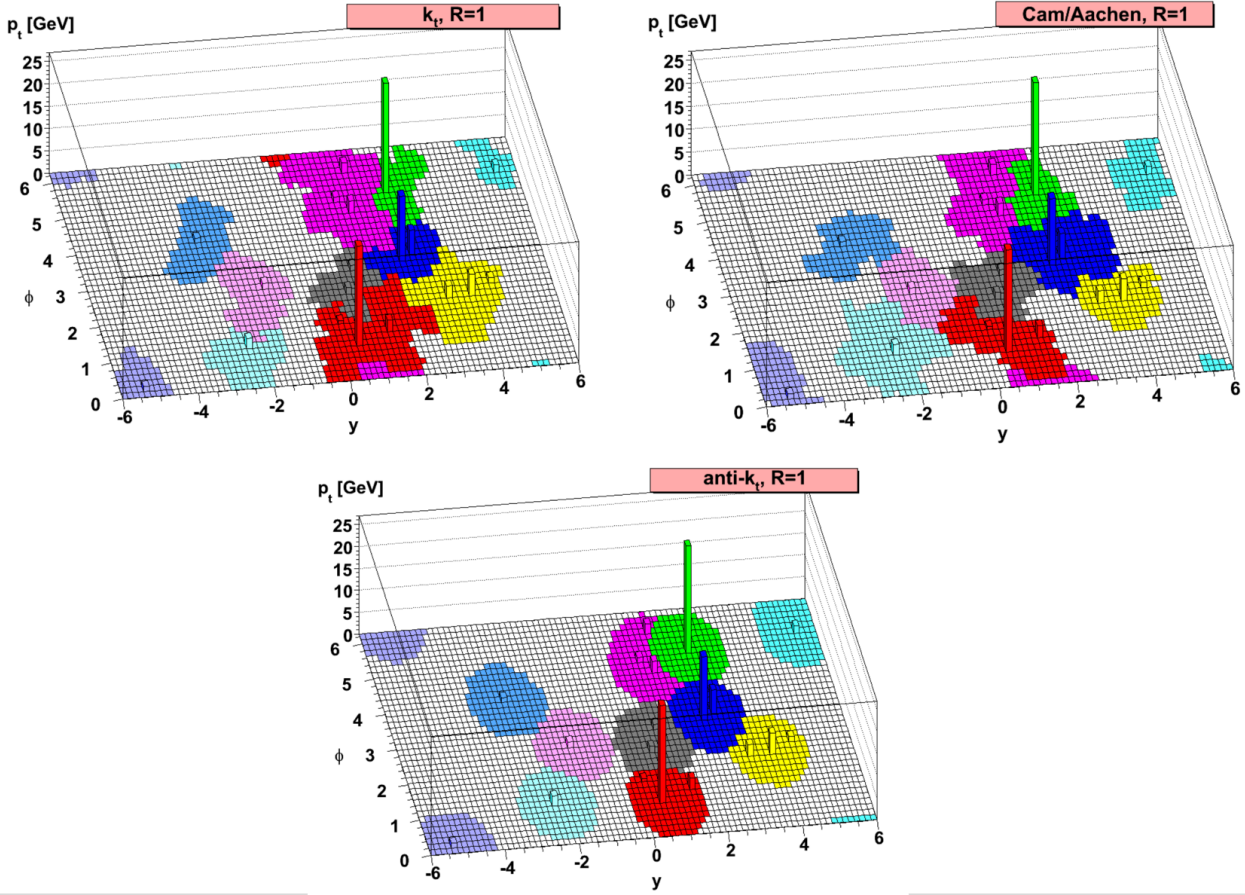


Figure 4.4: A sample parton-level event together with many random soft “ghosts”, clustered with three different jets algorithms, [70].

4.2 Inputs to jet reconstruction

Before we get into the details of the grooming techniques and the calibration procedures that need to be applied to jets, it is useful to go through the various inputs that can be used for jet reconstruction. Traditionally, when talking about jets in ATLAS, one refers to jets built from calorimeter energy deposits called topo-clusters, described in section 3.1.

However, tracks are also used in several aspects of jet reconstruction. For example, as of 2016, the use of tracks is required to derive the mass, see section 4.6, of topo-cluster jets. The tracks, however, only enter the jet reconstruction procedure via ghost-association. Ghost association is a technique that allows us to add tracks to a jet without changing its kinematics. If a track is contained within the radius of a calorimeter-based jet, its ‘ghost’ version is added to the clustering process. This ‘ghost’ track is none other than a track with the same direction but infinitesimally small p_T so that it doesn’t change the properties of the reconstructed jet, [72]. Ghost-associated tracks are also necessary reference objects to compute calibration uncertainties.

Jets are also commonly built from stable truth particles with a lifetime of at least 10 ps, resulting

in truth jets. Such jets are needed as a baseline for any resolution measurement and to determine calibration factors.

Finally jets can also be reconstructed from TCCs (section 3.3.1), PFOs (section 3.3.2) and UFOs (section 3.3.3) resulting in a dramatic improvement in terms of jet substructure resolution and, consequently, in our ability to correctly identify (tag) jets originating from vector bosons.

4.3 Jet grooming

As explained in the introduction, the environment surrounding a collision can be quite busy. On top of the hard-scatter event one has to take into account pile-up, the presence of initial and final state radiation (ISR and FSR) and the underlying event⁽²⁾ (UE). All these phenomena leave energy deposits in the detector that degrade the resolution of substructure variables. The removal of soft radiation originating from the underlying event and pile-up is known as *grooming* and can happen either during or after the process of iterative recombination. Several techniques based on different criteria can be used and for each of said methodologies many configurations can be applied. Only trimming [73] and soft-drop [74] are illustrated in the following as they are the ones used to groom the jets that are the object of the studies presented in this thesis.

4.3.1 Trimming

Jets reconstructed with the Anti- k_t algorithm have distinguishable substructure features depending on their nature, that is, whether they are generated by the decay of a massive vector boson accompanied by QCD showering or if they are produced by QCD alone. In the former case we expect two dense cores of high energy, whereas in the latter only one dense core should be present. To identify such particularities the constituents of the Anti- k_t large- R jets ($R = 1.0$) are reclustered into subjets by using the k_t algorithm with a radius parameter of $R_{sub} = 0.2$. Indeed, as already expressed in the previous section, such an algorithm is expected to reconstruct the QCD shower starting from low momentum particles and finishing with the hard scatter particles (involving larger p_T) [75].

A first, intuitive way of removing the underlying event, based on the characteristics of the jets, would be to apply a cut on the hardest N_{cut} subjets⁽³⁾. Said cut, however, can only work if its value is changed on the basis of the nature of the jet. Since trimming needs to work on both QCD only jets and vector boson jets this is not a viable solution. Instead, a cut on the p_T fraction of the subjet is used as a criterion to keep or discard the various constituents. The algorithm is illustrated in figure 4.5 and proceeds as follows:

- Jets are reconstructed with the Anti- k_t algorithm with a radius parameter R_{jet} ($R_{jet} = 1.0$).
- Subsequently, the constituents of the reconstructed jet are reclustered using the k_t algorithm with a radius parameter R_{sub} ($R_{sub} = 0.2$).

⁽²⁾In proton-proton collision there can be multiple interactions between the quarks and gluons that compose them. As such, we refer to the most energetic qq, qg or gg interaction as the hard scatter event and the remaining interactions as the underlying event.

⁽³⁾Subjets are jets reconstructed within the jet.

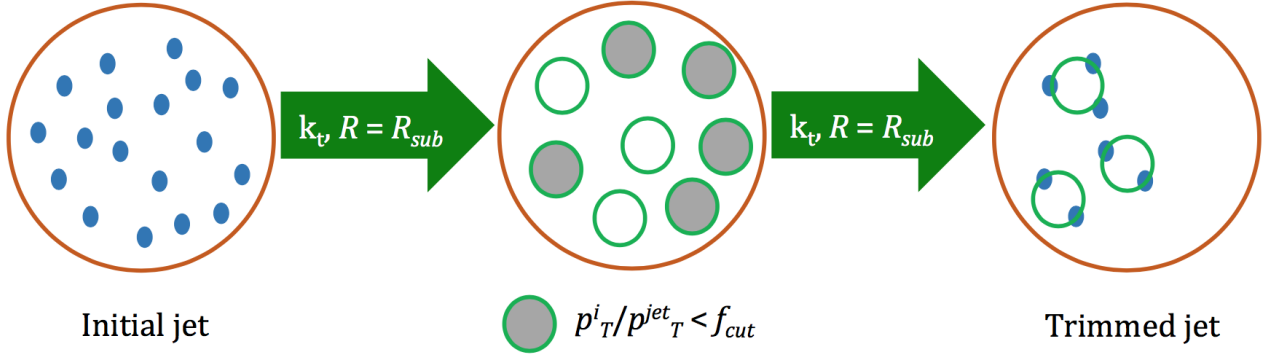


Figure 4.5: Diagram depicting the jet trimming procedure.

- A cut on the p_T fraction of the reconstructed subjets is applied. Any subjet i with $\frac{p_T^i}{p_T^{jet}} < f_{cut}$ is discarded, where f_{cut} is a parameter of the algorithm and is, in this case $f_{cut} = 5\%$.
- The remaining constituents form the trimmed jet.

The main goal of trimming is to exclude regions of the jet where the energy deposits come from the underlying event, however, as seen in Run 1, this technique also has a welcome secondary effect of reducing the impact of pile-up in large- R jets.

4.3.2 SoftDrop

Soft drop declustering, very much like trimming, was developed to reduce the impact of pile up, ISR and the underlying event by removing wide-angle soft radiation [74]. The method to achieve this result, however, differs. If we consider a jet of radius parameter R_0 with only two inputs, each with transverse momentum p_{Ti} and angular separation ΔR_{ij} , soft drop will remove the least energetic constituent unless the following condition is true:

$$\frac{\min p_{T1}, p_{T2}}{p_{T1} + p_{T2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta. \quad (4.3)$$

z_{cut} and β are, respectively, the soft drop threshold and the angular exponent and their numerical values control how much of the soft energy contributions are removed. By construction, the condition fails if the constituents are far from each other in the angular plane, if $\beta \rightarrow \infty$ the returned jet will be ungroomed [74]. The procedure described can be extended to jets reconstructed with the anti- k_t algorithm and containing more than two constituents. First the constituents of the jet need to be reclustered using the Cambridge-Aachen algorithm which returns a pairwise clustering tree with an angular ordering. The following steps are then applied iteratively:

- the jet is split into two subjets corresponding to the last stage of the C/A clustering.
- if the subjets pass the soft-drop condition described previously then the current jet is the final soft-drop jet.

- if the subjects fail the condition, the jet is redefined to be equal to the subject with the larger transverse momentum.
- in case the final stage of the C/A algorithm cannot be undone, i.e there is a single constituent in the jet, the procedure stops and the final jet is the jet with a single constituent.

Soft drop grooming has an advantage with respect to trimming. Indeed, using this procedure to groom the jets means that some jet substructure observables are calculable beyond leading-logarithm accuracy [65]. The soft drop grooming procedure used in this thesis is parametrised as follow: $\beta = 1.0$ and $z_{cut} = 0.1$.

4.4 Pile-up mitigation techniques

In addition to the grooming techniques illustrated previously, pile-up in jets can also be tackled at the constituents level by preprocessing them prior to the jet reconstruction procedure. These techniques are called pile-up mitigation algorithms and can be used singularly or in combination. The algorithms are applied on the entire set of inputs if the jet is reconstructed using topoclusters while if the inputs are particle flow objects then only neutral constituents are targeted. Indeed using tracker information for charged/combined constituents means that the primary vertex already provides a powerful method to reduce pile-up contributions.

4.4.1 Constituent subtraction

Constituent subtraction (CS) builds on area subtraction⁽⁴⁾ and extends it to be used at the components level. The CS procedure starts by adding massless particles, called 'ghosts', on top of the energy deposits left in the calorimeter by the event under study [77]. These ghost particles have a very low momentum $p_T^g = A_g \times \rho$ where $A_g = 0.01$ is the area of the ghosts in the $\eta - \phi$ plane and ρ is the energy pile-up density. ρ corresponds to the median of the $\frac{p_T}{A}$ distribution of the small radius ($R = 0.4$) jets, reconstructed with the k_t algorithm and with $|\eta| < 2.0$, in the event [65]. As such, ρ has to be computed for each event studied and, once obtained, one can derive the average pile-up contribution in the event.

After adding the ghost particles, the distance between each ghost i and cluster j is computed and the formed pairs are sorted in ascending order of $\Delta R_{i,j}$:

$$\Delta R_{i,j} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \quad (4.4)$$

⁽⁴⁾Area subtraction relies on the idea that, by suitably defining the area of a jet, the pile-up contamination will be proportional to said area [76].

The CS algorithm then runs through the ordered pairs of clusters and ghosts and modifies the p_T of the (i, j) pair by applying the following rules:

$$\begin{aligned}
 &\text{If } p_{T,i} \geq p_{T,j}: p_{T,i} \rightarrow p_{T,i} - p_{T,j}, \\
 &\quad p_{T,j} \rightarrow 0; \\
 &\text{otherwise: } p_{T,j} \rightarrow p_{T,j} - p_{T,i}, \\
 &\quad p_{T,i} \rightarrow 0.
 \end{aligned} \tag{4.5}$$

The procedure stops when the angular distance between the cluster and ghost reaches its maximum allowed value $\Delta R_{max} = 0.25$ ⁽⁵⁾, once this stage is reached, all the remaining ghosts are removed.

4.4.2 SoftKiller

The SoftKiller algorithm works in a simpler way, the idea is to remove soft constituents that do not meet the transverse momentum cut requirement. However, as explained in section 4.3, for such a method to work the cut has to be event based. The p_T^{cut} value is chosen so that the pile-up density ρ , described in the previous paragraph, is close to zero.

A grid in the $\eta - \phi$ plane, of length $l = 0.6$ ⁽⁶⁾, is applied to the event and the p_T cut value is chosen as the one that, after removing all the components that have a transverse momentum smaller than the cut, results in half of the grid squares being empty.

The pile-up mitigation strategy applied to the jets in this thesis is a combination of Constituent Subtraction followed by SoftKiller (CS+SK).

4.5 Jet calibration

Now that we have described jet reconstruction and grooming processes it is necessary to address the topic of calibration. Calibration is needed to account for detector effects such as the non compensating character of the calorimeter, section 2.4. The jet energy scale (JES) and jet mass scale (JMS) calibrations should ensure the correct measurement of the average energy and mass across the whole detector, [78]. To do so, Monte Carlo simulations are used so that the average reconstructed jet energy/mass corresponds to the truth level energy/mass, section 4.2.

Furthermore, systematic uncertainties have to be computed to account for the differences observed on the jet energy and mass scale between data and simulation by using well-known reference objects, such as track jets reconstructed from charged particles, [22].

4.5.1 Simulation based calibration

A truth jet is matched to a reconstructed jet if their angular distance satisfies: $\Delta R_{j_1 j_2} < 0.75 \cdot R_{jet}$. In general, the reconstructed jet energy response, defined as $R_E = E_{reco}/E_{truth}$, is smaller than one.

⁽⁵⁾This specific value is based on studies on small-R jet performances and is the one used for this study.

⁽⁶⁾This specific value is also based on studies on small-R jet performances and is the one used for this study.

This energy loss is due to both detector and reconstruction effects: inactive material, leakage due to particles passing through cracks in the calorimeter, calorimeter non compensation and signal losses due to noise threshold or inefficiencies of the reconstruction procedure. Also, the topo-clustering algorithm, briefly mentioned in section 4.2, influences the jet mass, and, therefore, its response $R_m = m_{reco}/m_{truth}$. Calibration accounts for such effects and corrects the jet energy and mass scale, from now on JES and JMS respectively. Calibration factors are derived from Monte Carlo simulations with the “numerical inversion” technique which relates the calorimeter’s energy response to the true jet energy, [78].

Since proper mass reconstruction is one of the main goals when studying large- R jets, mass calibrations also need to be derived. Since the mass is derived from the energy, the two quantities are highly related, however, mass is particularly susceptible to soft, wide-angle contributions that minimally affect the energy scale. The procedure measures the jet mass response after the JES calibration is applied, and then uses the same numerical inversion technique to derive the mass calibration factors. These factors are then applied to the reconstructed jet mass.

4.5.2 *In situ* energy calibration

Once the calibration factors are derived from MC, the differences between MC simulation and data need to be accounted for. The use of conservation of transverse momentum p_T in events in which a large- R jet recoils against a well-known reference object allows us to measure said differences. The following methods are used: first an η -intercalibration procedure is applied to correct the p_T of forward ($|\eta| > 0.8$) jets, then a Z +jet balance method, a γ +jet balance method and a multijet balance method are used to offer complementary coverage over a broad p_T range [78].

The η -intercalibration is a method that extends the calibration to the forward region of the detector ($0.8 < |\eta| < 2.5$). The calibration factors are derived from the p_T balance differences between a central reference jet and a forward jet in data and simulation.

The Z +jet balance method is used to derive *in situ* calibration factors for central ($|\eta| < 0.8$) jets. It uses the average momentum balance (R_{DB}) of the large- R jet (p_T^J) with respect to a leptonically decaying Z boson (p_T^{ref}):

$$R_{DB} = \left\langle \frac{p_T^J}{p_T^{ref}} \right\rangle \quad (4.6)$$

Any mismodelling in the jet energy scale can be evaluated using the double ratio of R_{DB} in data and simulation: if the double ratio is not equal to unity the deviation can be used as an *in situ* correction factor.

The γ +jet method follows the same procedure just described but uses γ +jets final states to extend the derivation of *in situ* correction factors to higher large- R jet p_T [78]. It relies on the ability to precisely measure the energy of photons with respect to that of large- R jets.

Together, the Z +jet and γ +jet techniques provide constraints on the jet energy scale for jets with p_T up to 1 TeV, for jets with an energy greater than that, correction factors are derived using the multijet balance method. This method takes advantage of events where the a high p_T large- R jet is balanced against a system that consists of multiple lower p_T small- R jets [78].

4.5.3 *In situ* jet mass calibration

Correction factors to account for the differences between data and simulation are also derived for the jet mass scale. The R_{trk} method relies on the independence of measurements of jets' properties in the tracker portion of the detector and in the calorimeter. Jets built from tracks only account for the energy deposition of charged-particle constituents whereas calibrated topo-cluster jets provide a measure of the properties of the full shower [78].

$$r_{trk} = \left\langle \frac{p_T^{calo}}{p_T^{track}} \right\rangle \quad (4.7)$$

The average r_{trk} response, that is the average ratio of transverse momentum in calorimeter jets and track jets, is proportional to the average ratio of transverse momentum in calorimeter jets and truth jets. The double ratio of the response in simulation and data ($R_{trk} = r_{trk}^{data}/r_{trk}^{MC}$) should therefore be equal to unity for well-measured observables. Any deviation can then be taken as a scale uncertainty in the measurement. This method can be used to determine uncertainties on a number of variables such as the mass and substructure information.

The calibration procedure described in this section is applied to calorimeter jets. In the analysis presented in chapter 5, TCC jets are used. In that case only the simulation based calibration is applied. Indeed, since the primary goal is to optimise the ability to distinguish QCD jets from vector bosons jets and not measure the jet four-vector, *in situ* correction factors are less relevant. Instead, the R_{trk} method is used to derive uncertainties on the transverse momentum of the jets used in the analysis.

4.6 Jet substructure variables

As already mentioned, a large- R jet originating from a massive vector boson will have significantly different features than a large- R jet with the same kinematics originating from a light quark or gluon (QCD jet). Such jet properties can be quantified by using so-called *Jet Substructure Variables* (JSS) or *moments*, that is, quantities derived from the constituents of the jets, regardless of whether they are topo-cluster 3.1, truth particles, track-calo clusters 3.3.1 or united flow objects 3.3.3. The following text describes the substructure moments commonly used by the ATLAS collaboration because of their strong ability to discriminate between the two-prong structure of vector boson jets and the one-prong structure of QCD jets, [22]. These substructure variables can also be used to distinguish between differently polarised vector bosons and their use will be explored in chapter 7. The sensitivity of the various JSS to polarisation can be seen by comparing the distributions of longitudinally and transversely polarised W bosons. The discrimination power of these variables, that is their ability to highlight the differences between jets of different nature or polarisation, inevitably depends on their resolutions, and, as such, on how the jet is reconstructed. For demonstrative purposes we will show the variable distributions for large- R jets ($R = 1.0$) reconstructed with the Anti- k_t algorithm, groomed with SoftDrop and applying the CS+SK procedure described earlier, where the inputs to jet

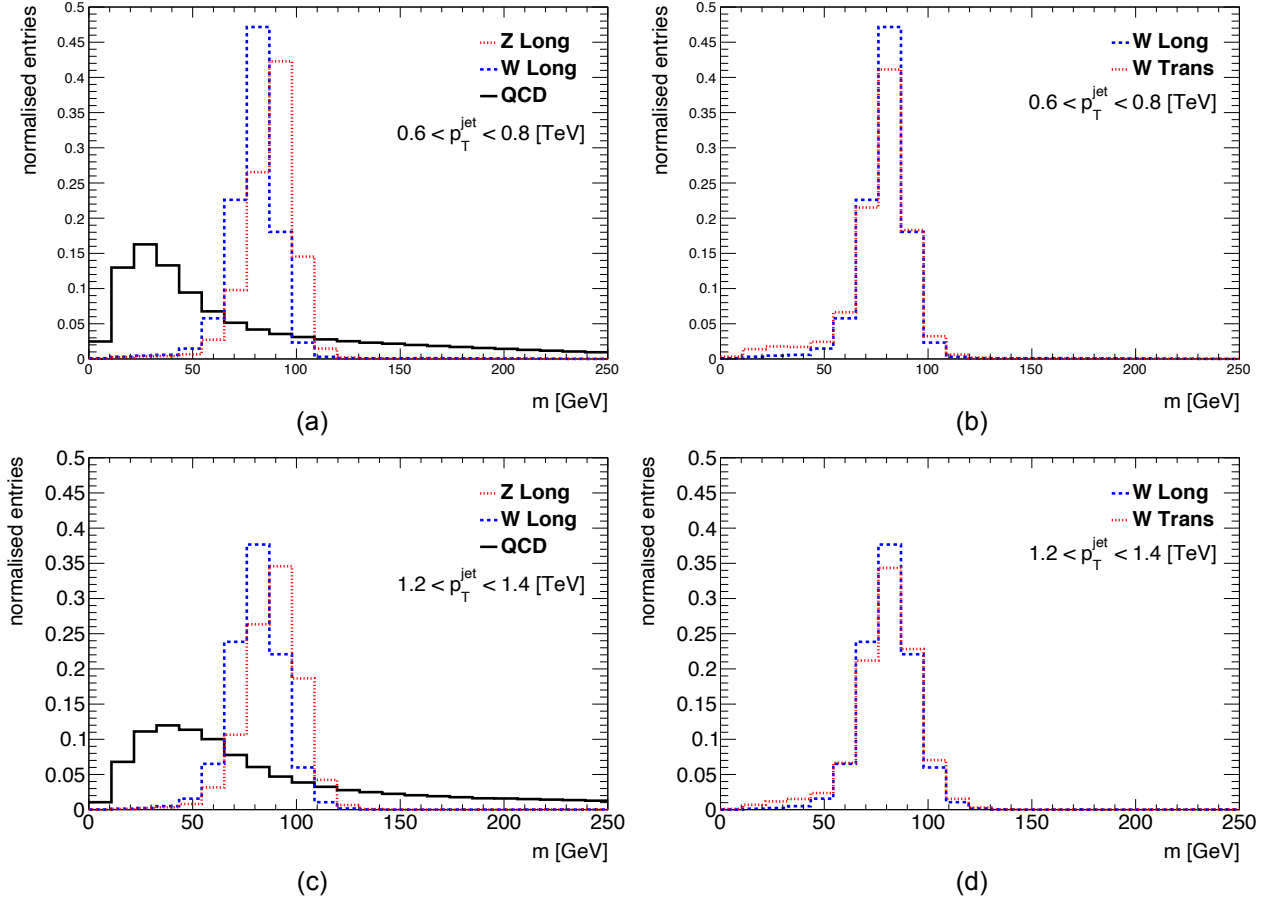


Figure 4.6: Mass distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

reconstruction are topo-clusters ⁽⁷⁾.

4.6.1 Mass

Mass is the most intuitive jet substructure variable and it is also a very powerful tool to distinguish a QCD jet from a vector boson jet and to identify the boson type (W or Z). Large- R jets originating from W/Z bosons have masses close to that of said bosons. On the other hand, background jets originate from nearly massless partons. Their mass, however, is significantly larger than the parton mass. This effect is tied to a mixture of factors, the main ones being the finite spatial resolution of the calorimeter and the stochastic nature of hadronic showers. The two types of distributions are clearly shown in

⁽⁷⁾Multijet production is used to define the QCD background originating from light quarks and gluons. These processes are generated using Pythia v8.230 [79], the NNPDF2.3LO PDF set [80] and the A14 set of tuned parameters [81]. For the two nominal W/Z boson signal samples, high-mass SSM $W' \rightarrow WZ \rightarrow qq\bar{q}\bar{q}$ events are simulated using Pythia v8.235 [79] with the NNPDF2.3LO PDF set [80] and the A14 set of tuned parameters [81] and filters are applied to select the polarisation of the vector bosons [82].

figure 4.6. For a jet J with constituents i , it is defined as follows:

$$m = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} \vec{p}_i\right)^2} \quad (4.8)$$

The definition is the same independent of the nature of the constituents. For calorimeter jets the mass, m_{calo} , is therefore derived from the topo-cluster constituents. However, in Run-2, the mass of calorimeter large- R jets has been redefined to make use of tracker information. This implementation has become necessary since the angular spread of the decay products of a massive particle scales like $1/p_T$, as described in the introduction of this chapter. In Run-2, the energies reached at the collision point are so large that massive vector bosons can be produced with a boost such that the angular spread between the two quarks is comparable with the calorimeter granularity, [83]. A new mass, the *combined mass* m_{comb} is therefore used for calorimeter large- R jets. It consists of the linear combination of the usual calorimeter jet mass m_{calo} , and the so-called track assisted jet mass, m_{TA} with weights defined by their respective mass resolutions σ .

$$m_{TA} = m_{tracks} \times \frac{p_T^{calo}}{p_T^{tracks}}$$

$$m_{comb} = m_{calo} \frac{\sigma_{m_{calo}}^{-2}}{\sigma_{m_{calo}}^{-2} + \sigma_{m_{TA}}^{-2}} + m_{TA} \frac{\sigma_{m_{TA}}^{-2}}{\sigma_{m_{calo}}^{-2} + \sigma_{m_{TA}}^{-2}}$$

where p_T^{calo} is the p_T of the calorimeter large- R jet and p_T^{tracks} and m_{tracks} are, respectively, the transverse momentum and the mass of the four-vector sum of the tracks associated to the large- R jet. The combined mass efficiently exploits the good energy resolution of the calorimeter at high p_T and the better p_T resolution of the tracker at low p_T . Since it is derived from m_{TA} and m_{calo} , which are both calibrated, and the sum of the weights is equal to one, the combined mass is also automatically calibrated.

4.6.2 Energy correlation variables

Many substructure moments are defined based on the properties and relationships of reconstructed subjects (like N-subjettiness also defined in this section), this, of course, makes the step of subject-finding necessary. Energy correlation variables, [84], are defined without the need of such additional step⁽⁸⁾ and, consequently, has a better treatment of soft emissions at large angles. Both features will be especially useful in the following of this thesis when developing jet tagging algorithms.

The first four N-point energy correlation functions (ECFs) for a jet J are defined as follow:

$$\begin{aligned} \text{ECF}(0, \beta) &= 1, \\ \text{ECF}(1, \beta) &= \sum_{i \in J} p_{T_i}, \\ \text{ECF}(2, \beta) &= \sum_{i < j \in J} p_{T_i} p_{T_j} (R_{ij})^\beta, \\ \text{ECF}(3, \beta) &= \sum_{i < j < k \in J} p_{T_i} p_{T_j} p_{T_k} (R_{ij} R_{jk} R_{ik})^\beta \end{aligned}$$

⁽⁸⁾Subject finding algorithms are applied anyway in the reconstruction of the jet for pruning purposes, as such there is no/little additional processing time needed.

where i, j, k are the constituents, R_{ij} is the euclidean distance between i, j in the pseudorapidity-azimuth angle plane and β is the angular exponent that has to be optimised based on the mass of the resonance studied, [84]. Since $(N + 1)$ point energy correlation functions (ECFs) are sensitive to N -prong substructure,[84], only these first ECFs are needed to distinguish the typical 2-prong substructure of W/Z jets from the 1-prong substructure of QCD jets.

From these ECFs, two dimensionless ratios, D_2^β and C_2^β , are defined and used in the discrimination of the two jets typologies of interest.

$$C_2^\beta = \frac{\text{ECF}(3, \beta) \text{ECF}(1, \beta)}{\text{ECF}(2, \beta)^2}$$

$$D_2^\beta = \frac{\text{ECF}(3, \beta) \text{ECF}(1, \beta)^3}{\text{ECF}(2, \beta)^3}$$

In the specific case of W,Z vector bosons decaying hadronically, the theoretically optimal value is $\beta = 2$, while $\beta = 1$ is found to be experimentally useful. Most of the studies in the rest of the thesis will be made on the $D_2^{\beta=1}$ variable, from now on simply referred to as D_2 . The different shapes of the D_2 distribution for different boson types can be seen in figure 4.7.

4.6.3 N-subjettiness

The τ_N N-subjettiness variables, [85], are related to the number of subjet axes within a jet. Their aim is to assess how compatible is the substructure of the large- R jet with the hypothesis of it containing N or fewer subjet axes.

The k_t reconstruction algorithm is applied to the large- R jet, J , in question and its k constituents, and is required to build exactly N subjets. Then τ_N is derived as follows:

$$\tau_N = \frac{1}{d_0} \sum_{k \in J} p_{T_k} \min \Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}$$

$$d_0 = \sum_{k \in J} p_{T_k} R_{jet}$$

where d_0 is a normalisation factor defined by the radius parameter R_{jet} and $\Delta R_{I,k} = \sqrt{(\Delta \eta_{I,k})^2 + (\Delta \phi_{I,k})^2}$ is the distance between the I^{th} subjet and the k^{th} constituent.

Small values ($\tau_N \approx 0$) of N-subjettiness indicate that most of the radiation contained within the large- R jet is aligned with the direction of the N subjets. Such large- R jets will therefore be well described by N or fewer subjets. Typically, ratios of such variable are studied as they have larger discriminating power: a jet with an N -prong substructure will have small values of the $\tau_{N,N-1} = \tau_N / \tau_{N-1}$ ratio.

In this thesis we will mostly focus on τ_1 , τ_2 and their ratio τ_{21} , their discriminating power can be seen in figure 4.8.

4.6.4 Other substructure variables

Split12 Split12 (see figure 4.9) is the distance of the last clustering step when reconstructing the jet with the k_t algorithm, that is the equation 4.2 with parameter $p = 1$.

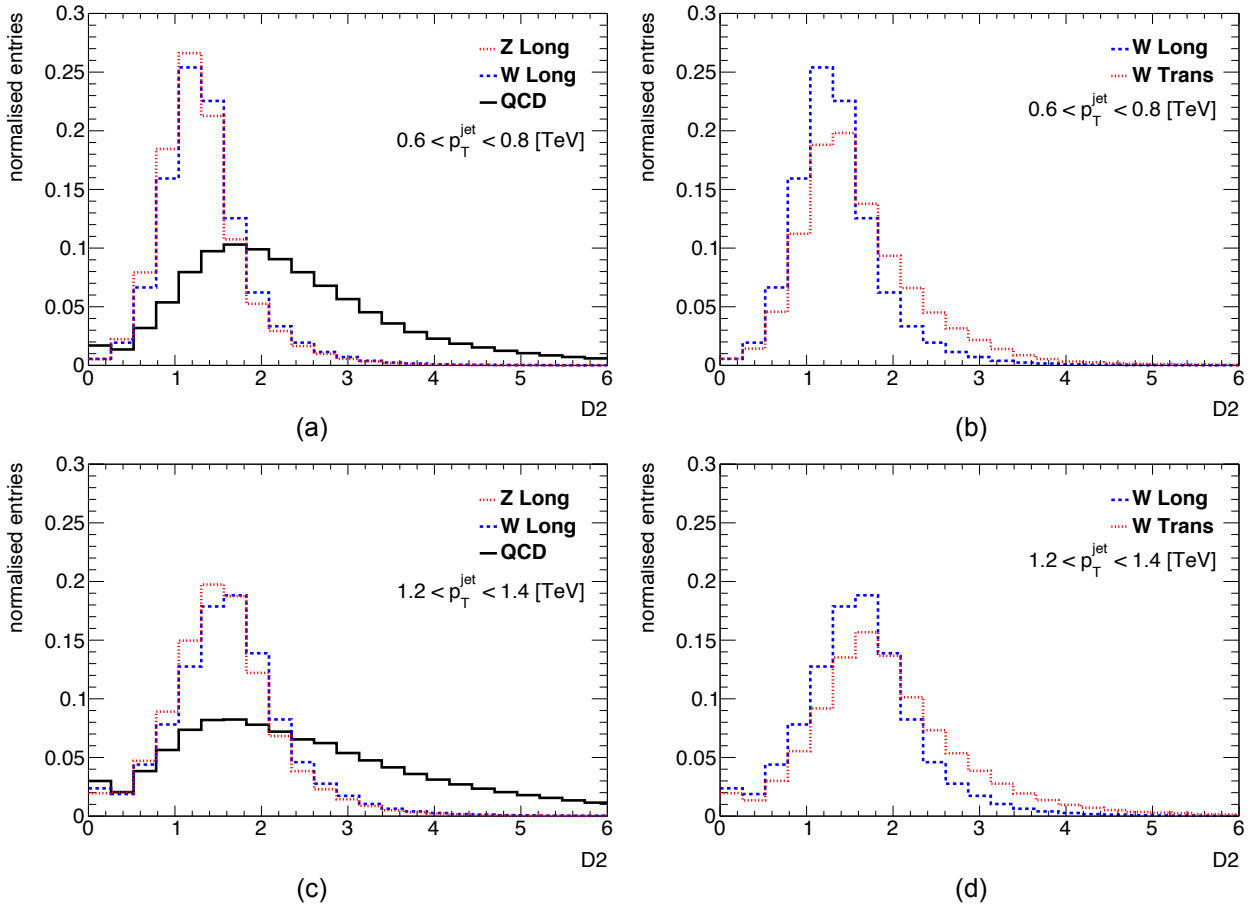


Figure 4.7: D_2 distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

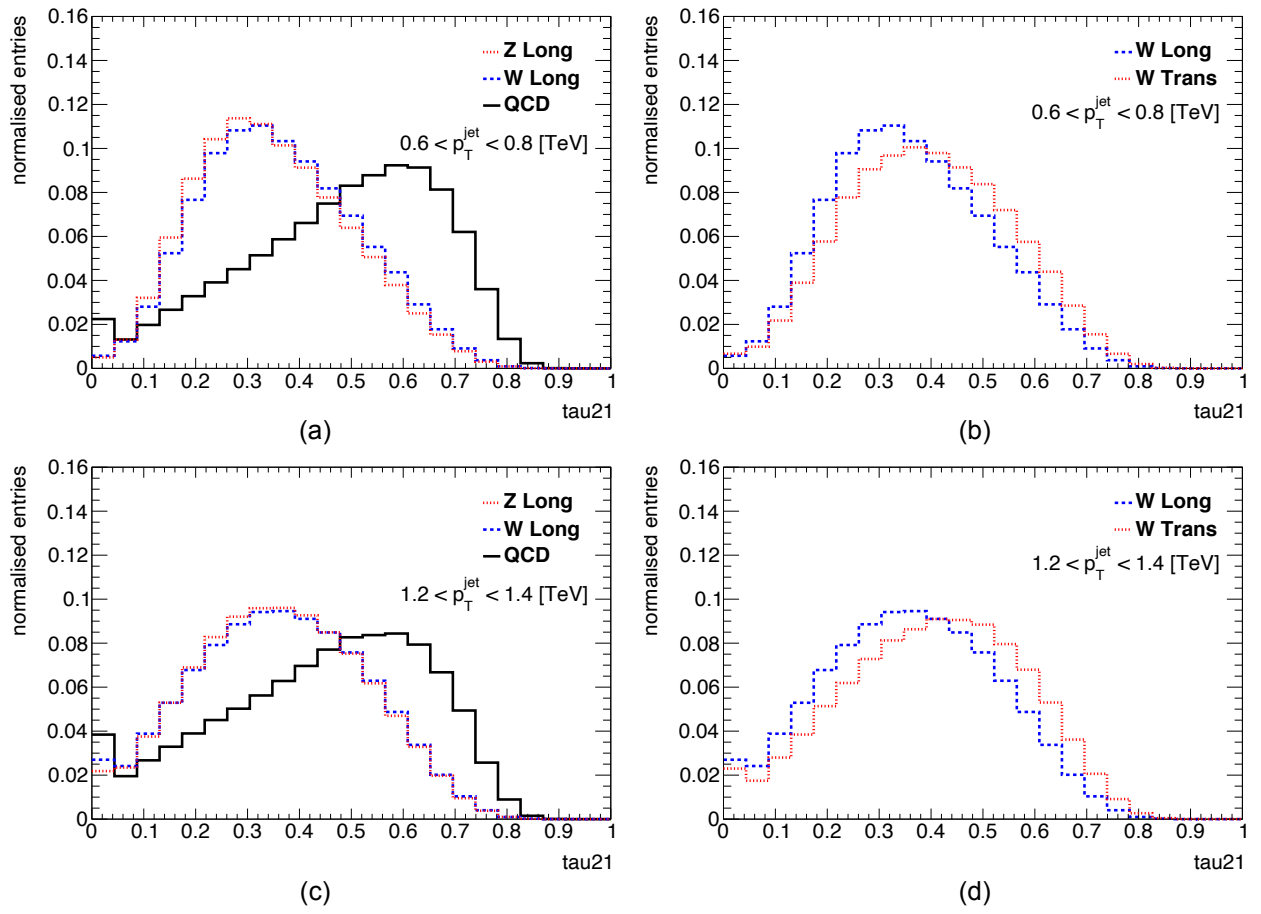


Figure 4.8: τ_{21} distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

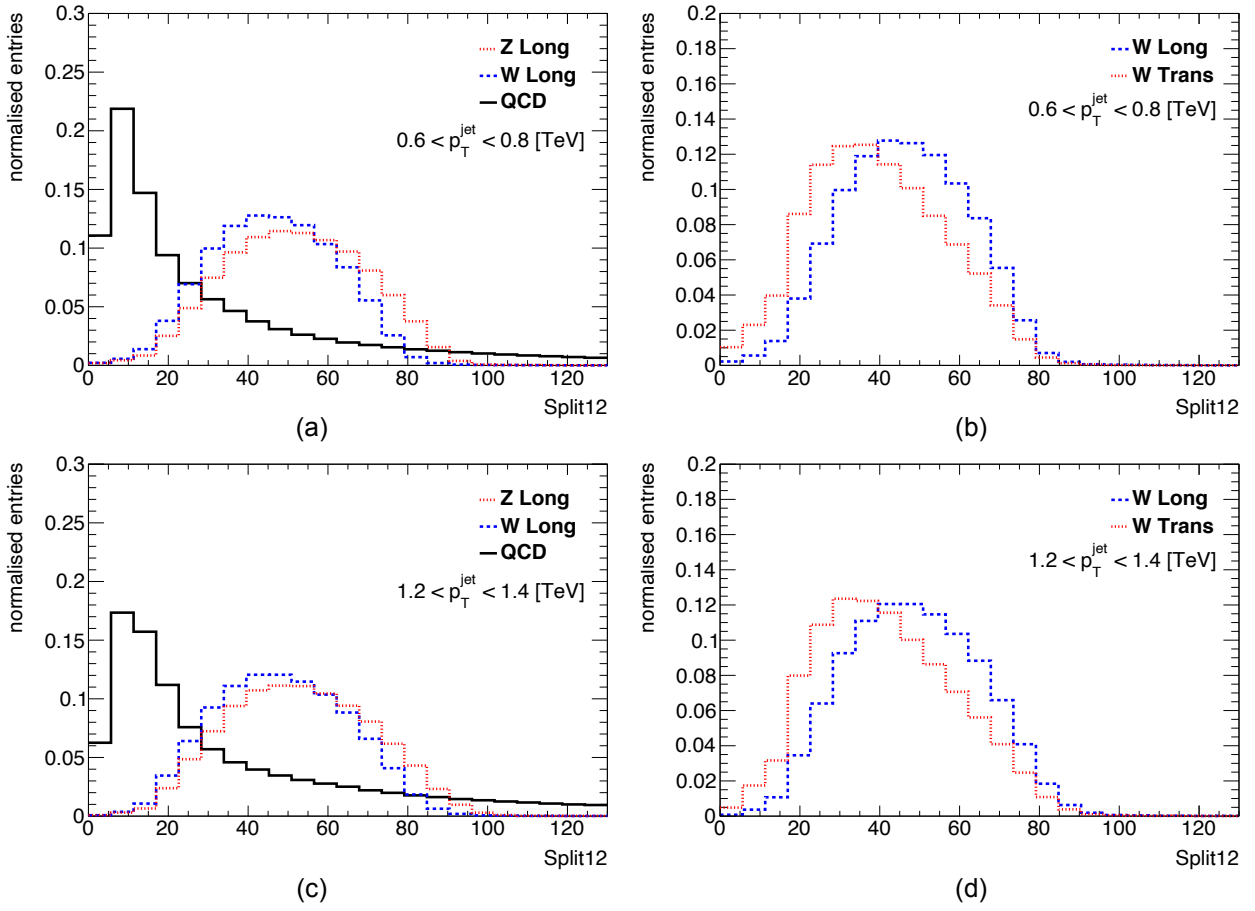


Figure 4.9: Split12 distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

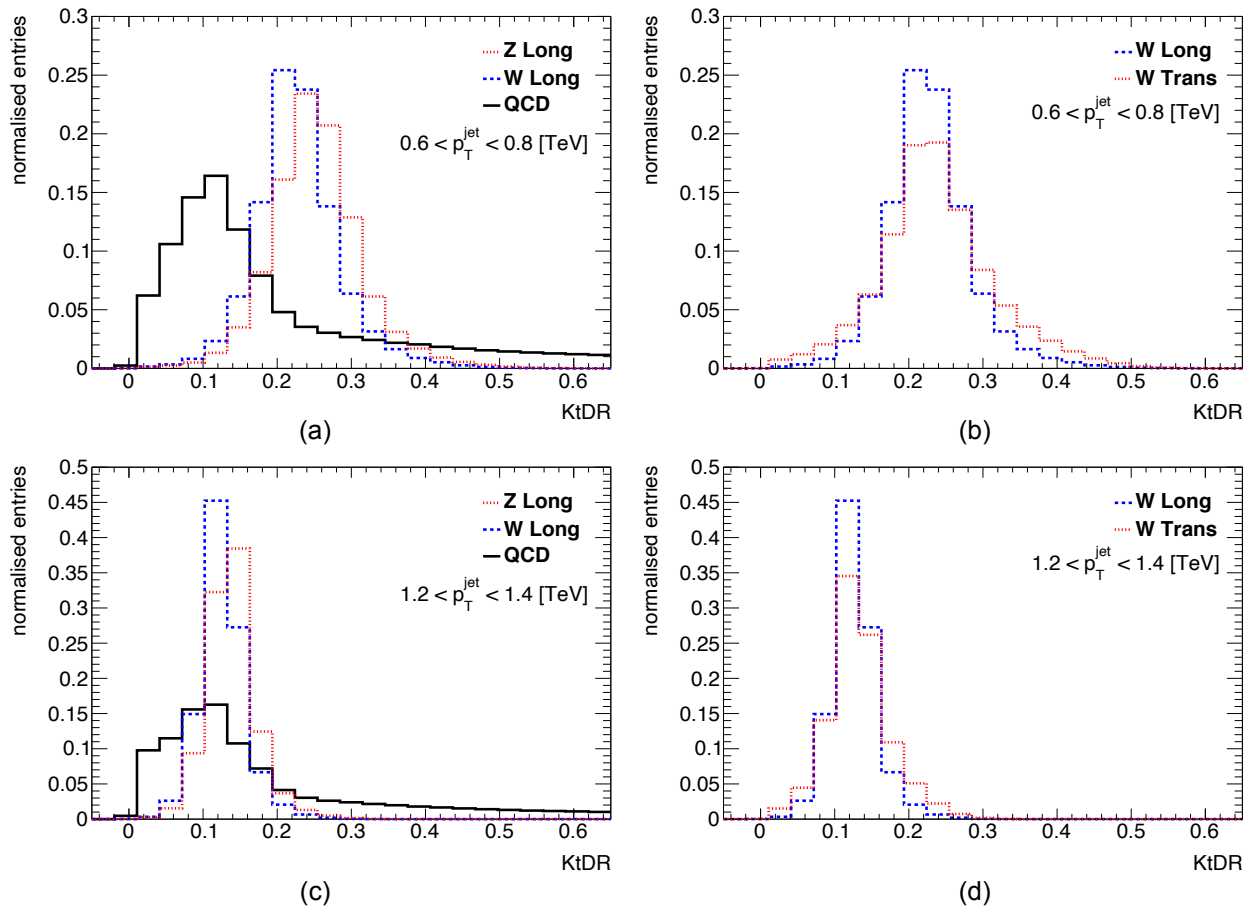


Figure 4.10: KtDR distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

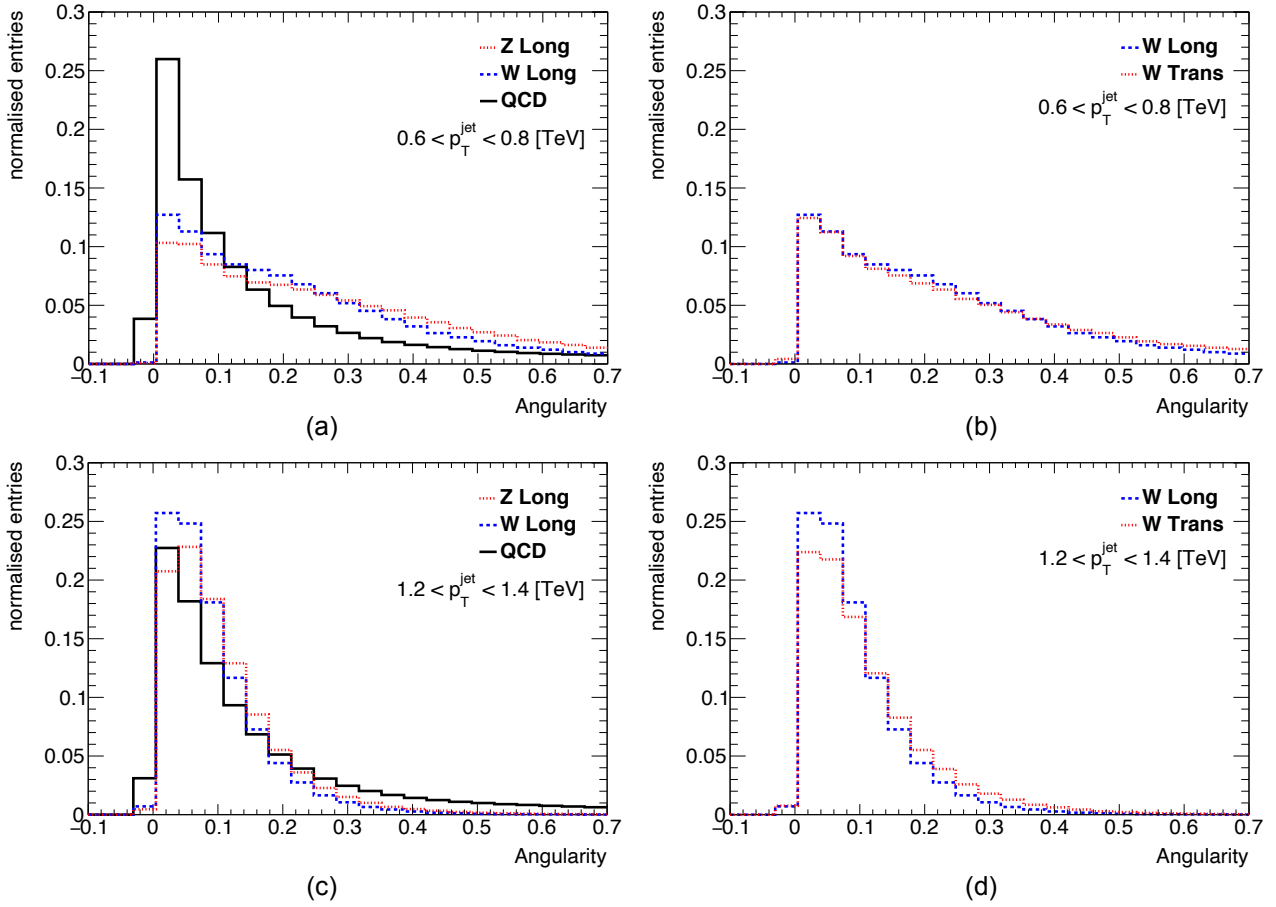


Figure 4.11: Angularity distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

KtDR KtDR, shown in figure 4.10, is none other than the ΔR used in the k_t algorithm and defined in equation 4.2, in particular KtDR refers to the angular distance in the final k_t clustering step.

Angularity Angularity relates to how symmetric the energy distribution is within a jet. It is generally defined as:

$$a_\beta = \frac{1}{M} \sum E_i \sin^\beta \theta_i [1 - \cos \theta_i]^{1-\beta} \quad (4.9)$$

Where M is the mass of the jet and E_i and θ_i are the energy of the constituent and its angle with respect to the jet axis respectively. The β parameter determines if more weight is given to energy deposits close to the edges of the jet ($\beta < 0$) or to energy deposits close to the core of the jet ($\beta > 0$), [86]. In particular, for the purpose of this thesis we have that $\beta = -2$, this is indeed the value that is most sensitive to differences between QCD and vector bosons. The angularity distribution is shown in figure 4.11.

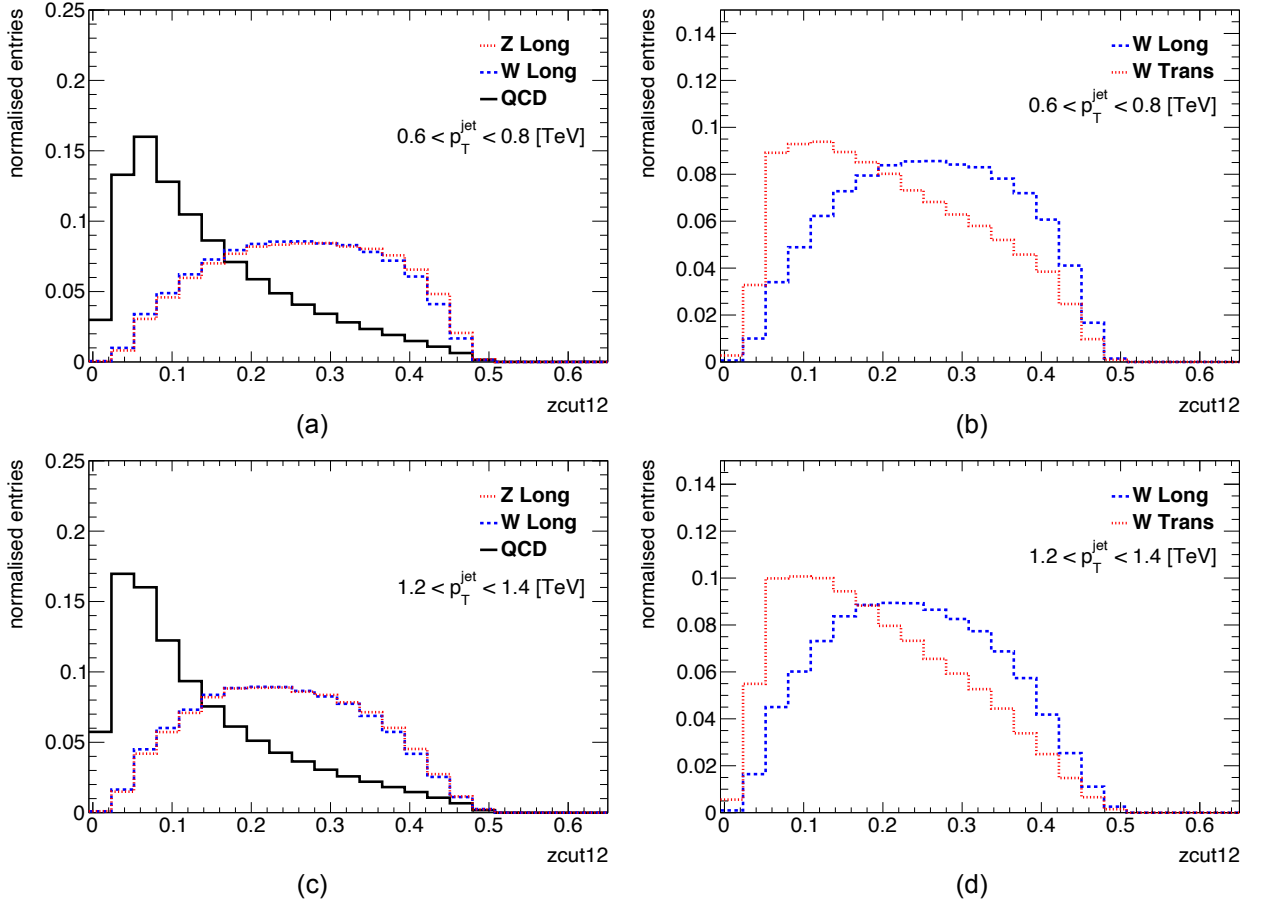


Figure 4.12: Zcut distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

Z_{12}^{cut} The Z_{12}^{cut} variable (see figure 4.12), not to be confused with z_{cut} defined for the SoftDrop pruning procedure, is a variable that quantifies the energy sharing between the first two splittings of the hadronic shower of a given parent particle. In the case of a vector boson decaying hadronically the first two splittings should roughly correspond to the two hadrons and should be balanced in energy whereas for QCD jets the energy splitting should be random. We have:

$$Z_{12}^{\text{cut}} = \frac{d_{12}}{d_{12} + m_{\text{jet}}} \quad (4.10)$$

where d_{12} corresponds to the splitting scale of the last step of the k_t clustering, see section 4.1, and m_{jet} is the mass of the jet, [87].

Aplanarity The Aplanarity is defined as $A = \frac{3\lambda_3}{2}$ where λ_3 is one of the eigenvalues that come out of the diagonalisation of the Sphericity tensor:

$$S^{k,l} = \frac{\sum p_i^k p_i^l}{\sum |\vec{p}_i|^2}, \quad (4.11)$$

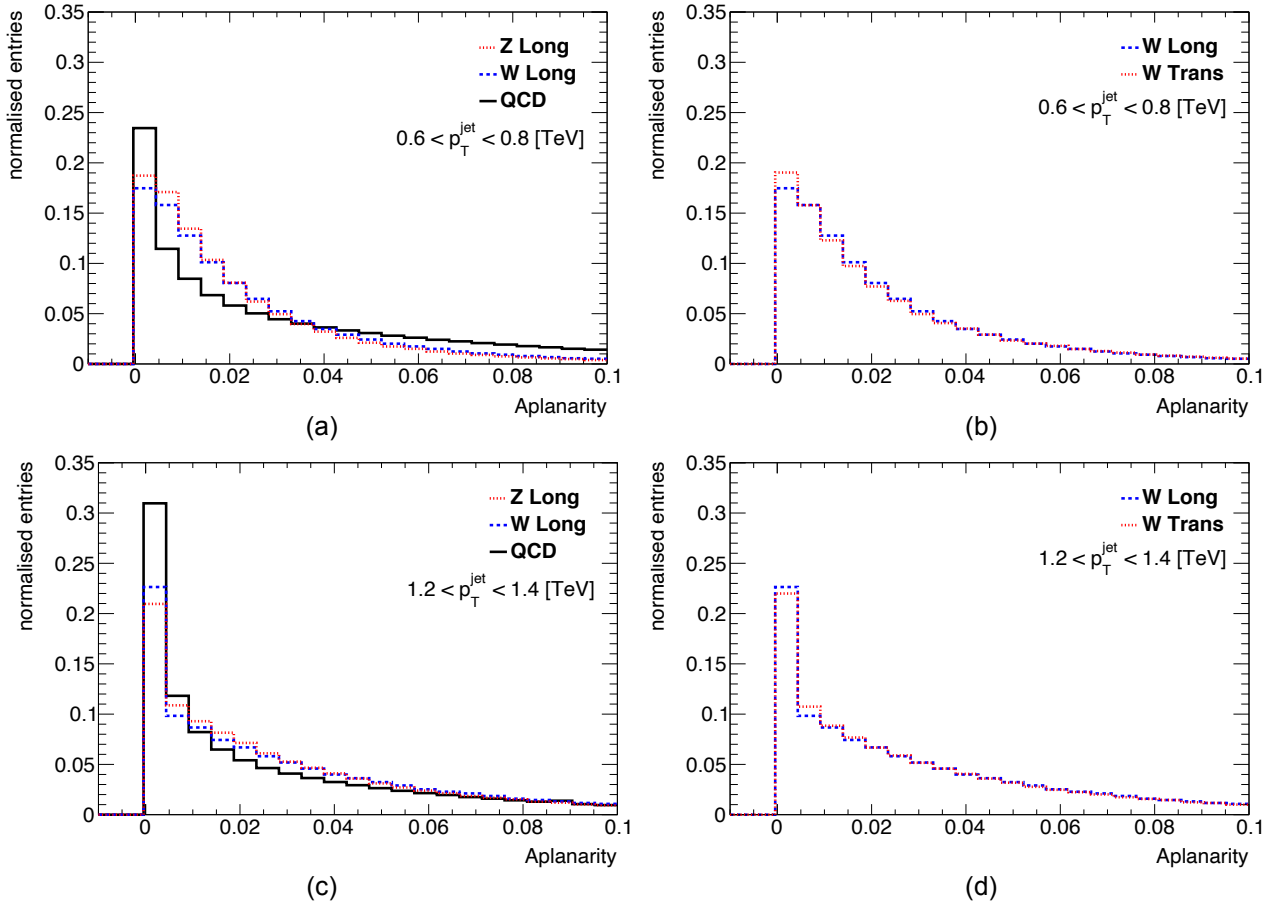


Figure 4.13: Aplanarity distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

where p_i^k is the k^{th} component of the transverse momentum relative to the jet axis. Naively, an aplanarity value of 0 corresponds to a highly directional distribution of the components whereas a value of 0.5 describes an isotropic distribution of the energy deposits, [88]. The distribution of the aplanarity variable can be seen in figure 4.13

R_2 R_2 is the ratio of the second-order and zeroth-order Fox-Wolfram moments. These moments are derived from Legendre polynomials, $P_l(x)$ and are defined as follows:

$$H_l = \sum \frac{|\vec{p}_i| |\vec{p}_j|}{E^2} P_l(\cos \theta_{ij}). \quad (4.12)$$

Again, to give an intuitive explanation of what R_2 represents, let's consider how it would look for a jet with a two prong structure. In that case we have that $H_0 = 1$ and $H_2 \approx 1$, as such jet's coming from vector bosons will have an R_2 distribution peaking at values close to 1 whereas jets coming from QCD will generally have lower R_2 values as can be seen in figure 4.14, [88].

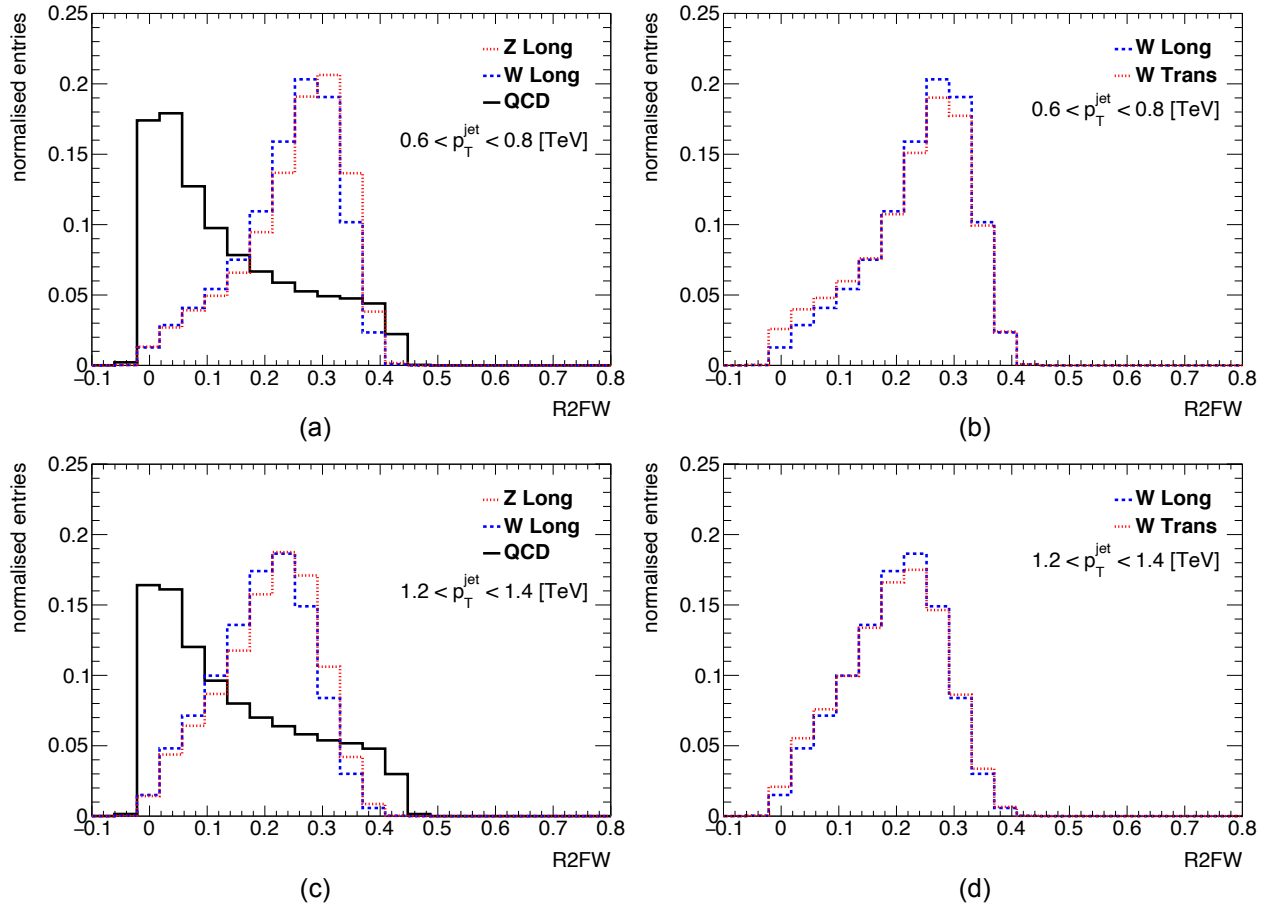


Figure 4.14: R_2 distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

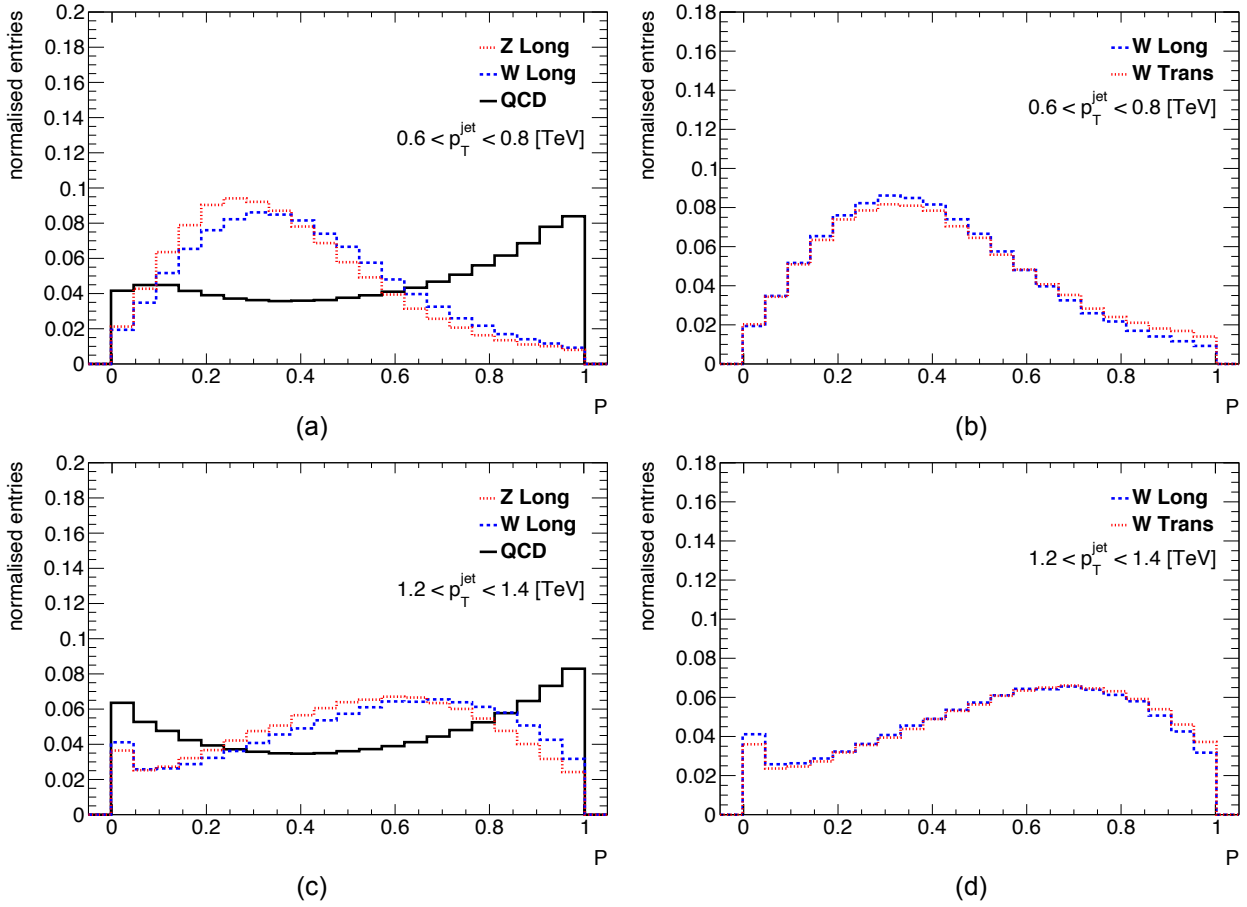


Figure 4.15: Planar flow distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

Planar flow Planar flow (see figure 4.15) is an indication of whether the energy is evenly spread over the plane across the face of the jet or linearly across the face of the jet, [86]. It is defined as follows:

$$Pf = \frac{4\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2} \quad (4.13)$$

where $\lambda_{1,2}$ are the eigenvalues of the energy shape tensor $I_w^{kl} = \sum \frac{p_{i,k}p_{i,l}}{w_i}$ where w_i is the energy of the jet component, and the k, l subscripts indicate the $k^{\text{th}}, l^{\text{th}}$ components of the transverse momentum relative to the jet axis. Pf is defined so that it approaches zero for objects with a linear energy flow (such as QCD) and unity if the energy deposits are isotropic in the jet, [89].

PtImbalance To try and observe differences in polarisation PtImbalance was defined. PtImbalance is simply the ratio $\frac{p_T^{\text{sub1}}}{p_T^{\text{sub0}}}$ where the 0 and 1 subscripts refer to the leading and subleading subjets respectively. To be defined, we need to be able to reconstruct at least two subjets ($R = 0.2$, k_t algorithm) within the large-R jet ($R = 1.0$, Anti- k_t algorithm), when jets have a large momentum this

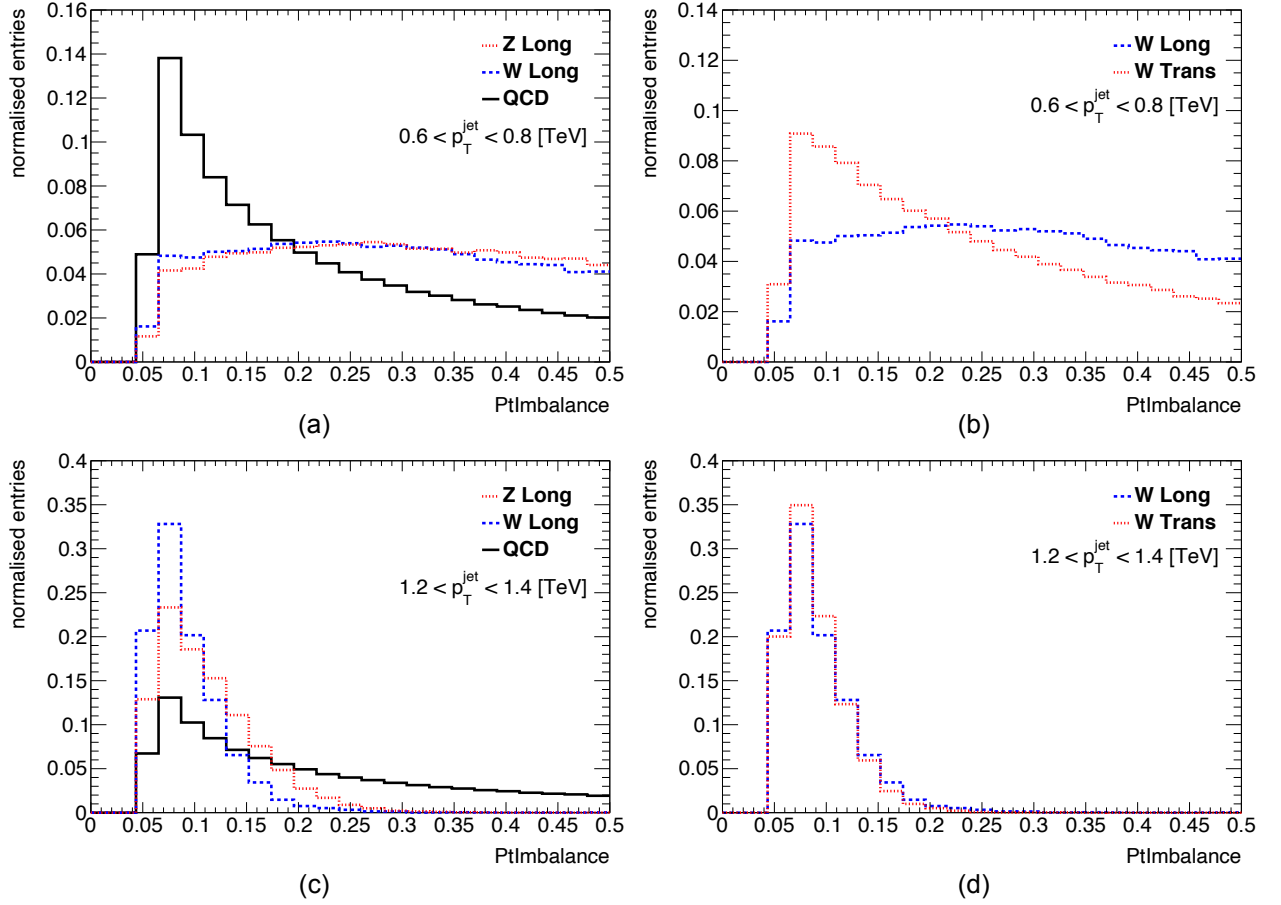


Figure 4.16: $PtImbalance$ distribution for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

condition is often not met, in these cases $PtImbalance$ is set to an arbitrary value of -0.1 ⁽⁹⁾. The variable distribution can be seen in figure 4.16.

Jet properties - number of tracks and number of constituents On top of the substructure variables described above, properties of the jet can also be used to increase our ability to correctly identify jets. In particular, for discriminating between W, Z and QCD, the number of tracks associated to the ungroomed jet, ug_{nTrk} and the number of constituents that compose the final jet, $nConst$ can be used. The distribution for ug_{nTrk} is shown in figure 4.17.

Several other substructure variables exist and can be used depending on the reconstruction algorithm used and the type of jet that needs to be studied. The elements that are key to this analysis, however, have been defined and described. Even though many of these variables have discriminating power, the use of a combination of these might not translate to a significant improvement in our jet

⁽⁹⁾The variable is positive by construction, as such any negative value will be enough to identify the jets with an ill-defined variable.

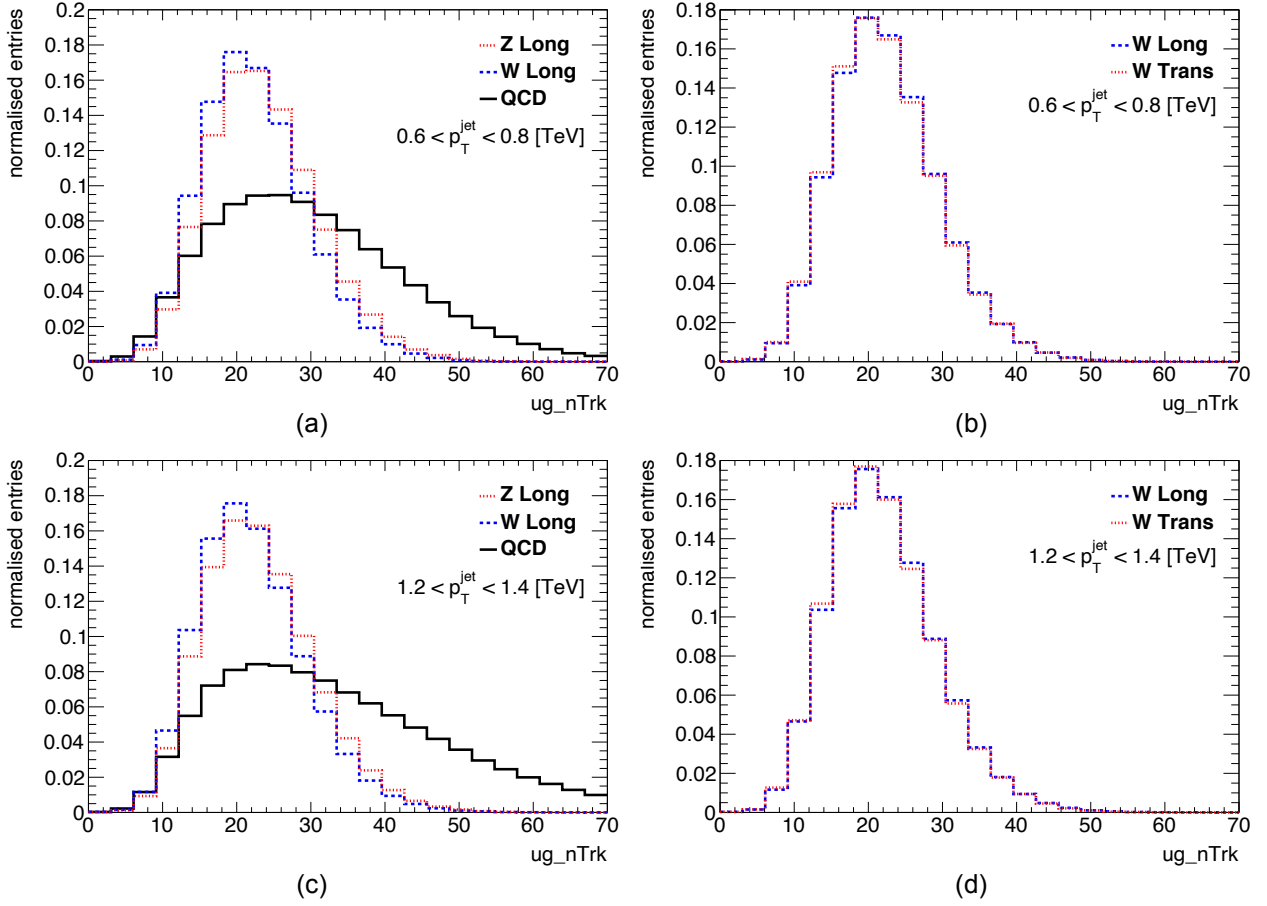


Figure 4.17: Distribution of the number of tracks associated to the ungroomed jet for (a, c) longitudinal Zs, Ws and QCD and (b, d) longitudinal and transverse Ws at (a, b) low p_T and (c, d) high p_T . The y axis represents arbitrary values.

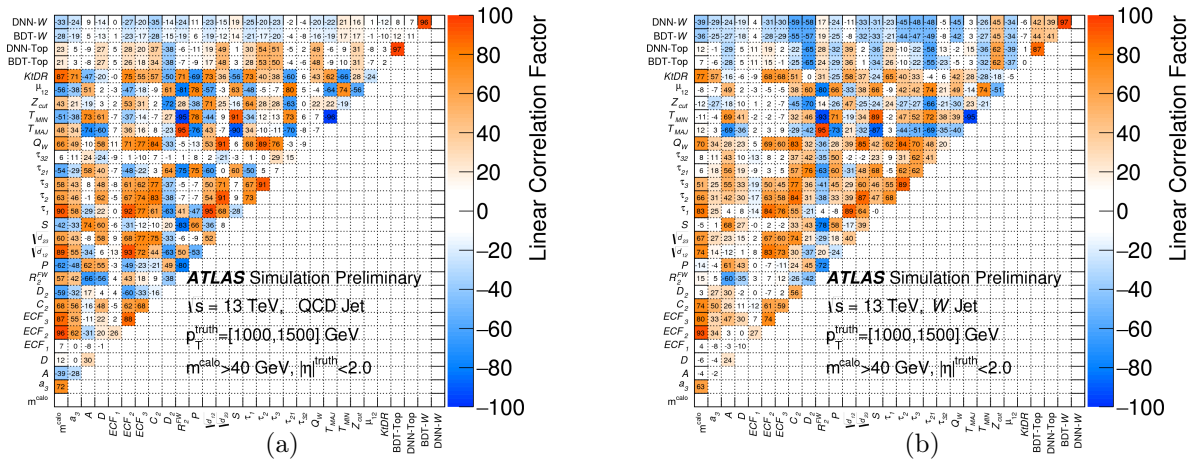


Figure 4.18: Linear correlation coefficient matrices for QCD in (a), and for W bosons in (b).

tagging abilities. The larger performance improvements will be seen by combining variables with little correlation amongst them or with correlations that differ between signal and background jets (see figure 4.18). The combination of highly correlated variables can still lead to small gains in performance and this phenomenon becomes particularly interesting in the context of Machine Learning algorithms (see chapter 6) which can take a large number of JSS variables as input: small gains for a large number of variables can add up to significant ones.

Part III

Searching for Resonances with Boson-Tagged Jets

Search for diboson resonances in hadronic final states

Contents

| | | |
|------------|---|------------|
| 5.1 | Data and simulated samples | 88 |
| 5.2 | Use of Track-CaloClusters for large-R jet reconstruction | 89 |
| 5.3 | Vector Boson Identification | 90 |
| 5.3.1 | Choice of significance metric | 91 |
| 5.3.2 | Choice of substructure variables and inputs to jet reconstruction | 91 |
| 5.3.3 | Performance evaluation and parametrisation of optimised tagger | 94 |
| 5.3.4 | Additional event selections applied in the analysis | 95 |
| 5.4 | Boson tagging efficiency | 98 |
| 5.5 | Event selection efficiency | 100 |
| 5.6 | Background parametrisation | 102 |
| 5.7 | Systematic uncertainties | 102 |
| 5.8 | Results | 105 |

All the reconstruction procedures described in the previous chapters are the building blocks of analyses at the ATLAS detector. They are used for physics measurements and resonance searches. The work presented in this thesis focuses on the latter: we are looking for a heavy narrow resonance decaying into a pair of boosted vector bosons ($X \rightarrow VV$ where V can be a W or a Z boson) that, in turn, decay in the fully hadronic mode ($V \rightarrow qq$). Such a signature could be an indication of new, beyond the standard model physics and is predicted by several SM extensions (as explained in section 1.4.1).

The choice of looking into the fully hadronic channel is supported by the large branching ratios of fully hadronic decays for W and Z bosons (respectively 67% and 70%). However these processes suffer from a large background contamination due to the production of multijet events.

The vector bosons (V) resulting from the decay of the searched-for resonance, given its large mass, are boosted ($p_T > 200$ GeV). This means, as discussed in section 4, that each V is reconstructed as

a single large- R ($R = 1.0$) jet J that contains both quarks produced during its decay: $V \rightarrow J$. We therefore often refer to this analysis as the VVJJ analysis.

The VVJJ analysis looks for a resonant peak in the dijet invariant mass spectrum. Previous analyses from ATLAS and CMS [90, 31], performed over the data collected in 2015 and 2016, showed no significant deviation from the smoothly falling QCD background. Searches over the same dataset, looking for a similar resonance in the semi-leptonic mode ($VV \rightarrow J l \nu$) [91], which benefits from less background contamination, also didn't find supporting evidence to the existence of such a hypothetical particle.

In this chapter we will describe the VVJJ analysis as performed on the data collected during the years from 2015 to 2018. On top of using a larger dataset, new techniques are introduced to increase the sensitivity of the search.

5.1 Data and simulated samples

The search is done over the data collected during the years going from 2015 to 2018, this period of time is referred to as Run2. The amount of data collected over this period of time, and that satisfies both the trigger conditions and quality requirements, amounts to 139 fb^{-1} . Out of all the collisions that happen at core of ATLAS only a selection is recorded and analysed. For the analysis presented here the following conditions are required:

- Events need to satisfy a single-jet trigger requirement. This means that at least one jet has to be reconstructed at each trigger level.
- At the high-level trigger, at least one topo-cluster large- R jet reconstructed with the anti- k_t algorithm with a transverse momentum larger than a certain threshold is required. Over the years 2015, 2016 and 2017-2018 the threshold was 360 GeV, 420 GeV and 440 GeV respectively.

While the analysis is performed over real data, the interpretation of the results and the development of the novel techniques presented in the rest of the chapter require the use of Monte Carlo simulated samples. Indeed, if a resonance were observed, signal samples produced in accordance with various theory scenarios would allow us to evaluate which theoretical interpretation is more plausible. On the other hand, if no particle is observed, the same signal samples allow us to set exclusion limits, that is to define our level of certainty with which we can exclude the existence of the particle under study. In addition, simulated signal samples allow us to develop tagging algorithms with which we can separate boson jets from QCD jets. Not all models that would fit our signature are used as inputs to the MonteCarlo generator, indeed, the branching ratios for the VV decay of the hypothetical particle need to be significant enough to be observed. With this in mind, the signal samples used in this analysis are created using three benchmark scenarios briefly described below, more details are provided in section 1.4.1.

The first scenario accounts for the existence of a massive spin 0 radion in an extension of the Randall-Sundrum framework [35, 36], this radion can decay into a pair of W s or a pair of Z s. The second scenario is based on two models, A and B, of the HVT phenomenological Lagrangian [38], which introduces a new spin 1 heavy vector triplet, Z' and W'^{\pm} , where Z' can decay into a WW pair and the two W' can decay into a W and a Z boson. Finally the third scenario uses the spin 2 bulk RS model [37] in which a Kaluza-Klein graviton, G_{KK} , can decay into two vector bosons: $G_{KK} \rightarrow WW$ or ZZ .

All three hypothetical particles described above also have other decay modes that are not described here but are accounted for in the generation of the samples.

Once the theoretical models are chosen, they can be used as inputs to Monte Carlo (MC) generators. MC samples for the radion, HVT, and RS models were generated using MadGraph 2.2.2 [92] interfaced to Pythia 8.186 [79] for hadronisation using the leading-order (LO) NNPDF 2.3 parton distribution function (PDF) [80] set and the ATLAS A14 set of tuned parameters [81] for the underlying event. In all signal samples, the W and Z bosons are primarily longitudinally polarised. After this first step the final-state particles (that is all the particles remaining at the end of the decay chain) are propagated through a detailed detector simulation based on GEANT4 [32].

To develop the boson tagger described in section 5.3 a dedicated sample of Sequential Standard Model (SSM) [93] W' decaying only into W/Z bosons that in turn decay hadronically is used. Finally, Pythia 8.186 [79] with the NNPDF 2.3 LO PDF [80] set and the A14 set of tuned parameters [81] was used to generate and shower multijet background events.

5.2 Use of Track-CaloClusters for large-R jet reconstruction

As described at the beginning of the Jet Reconstruction chapter 4, the decay products of boosted bosons are highly collimated. Due to this, the two quarks produced in the decay of a vector boson V are reconstructed within a single large-R jet. At the LHC we reach energies so high that the granularity of the ATLAS calorimeter isn't sufficient to separate the energy depositions originating from the two different quarks. This has direct consequences on the resolution of jet substructure variables and, therefore, on our ability to identify W s and Z s and separate them from the large QCD background. Tracker information can be used to compensate this lack of calorimeter granularity. Indeed new inputs to jet reconstruction called Track-CaloClusters and described at length in section 3.3.1 have been developed for this purpose. TCCs use the complementary behaviour of the tracker and the calorimeter at high transverse momentum.

All jets used in the VVJJ analysis are therefore built using TCCs as inputs. The anti- k_T jet reconstruction algorithm is used with a radius parameter $R = 1.0$ and a standard trimming procedure is applied (see section 4.3) to remove any subjet ($R = 0.2$ with k_T algorithm) with less than 5% of the p_T of the associated large-R jet [32].

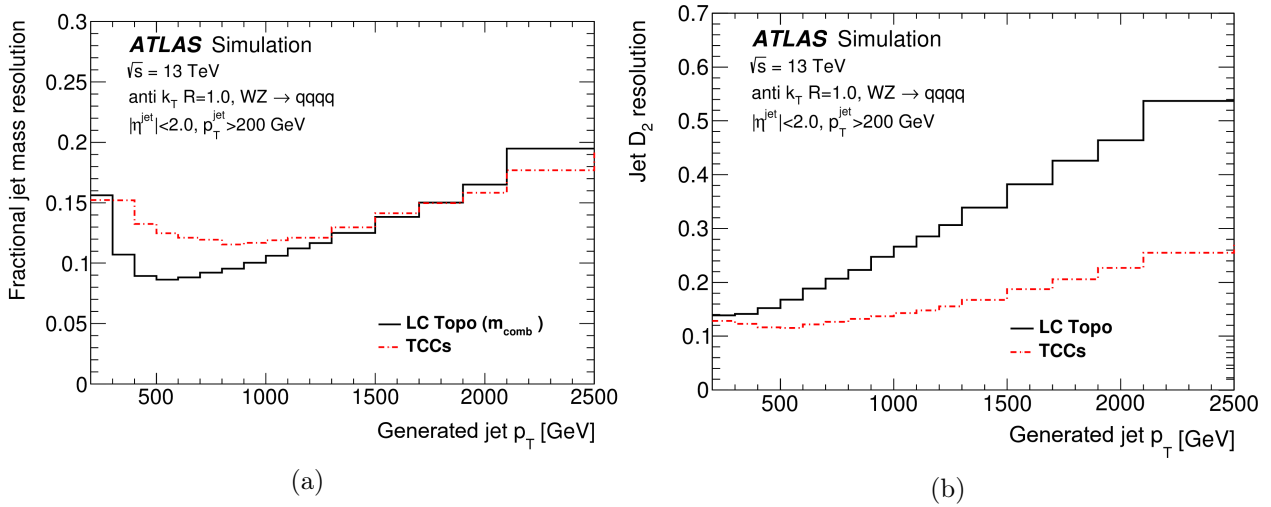


Figure 5.1: In (a), fractional jet mass resolution for jets reconstructed using topo-clusters in black and TCCs in red. For the topo-clusters jets the mass is a combined mass that uses track information to improve the resolution. In (b), the residual jet D_2 resolution for jets reconstructed using topo-clusters in black and TCCs in red [32].

The impact of the use of TCCs instead of topo-clusters as inputs to jet reconstruction can be seen in figure 5.1. The resolution⁽¹⁾ of the D_2 variable is drastically improved (roughly 2 times better at 2 TeV). The mass resolution instead is degraded below 1 TeV and outperforms the combined mass (see section 4.6) only above 2 TeV. As we will see the next section 5.3 the improvement on D_2 drastically outweighs the degradation in mass.

5.3 Vector Boson Identification

In the VVJJ analysis the multijet background is much larger with respect to the signature of interest. We therefore need a method to enhance the separation between the background contamination and signal of interest to have a chance of observing an excess over the smoothly falling background distribution. In the VVJJ analysis, this method is called jet tagging. A tagger's aim is to distinguish between multi-jet events and signal events. Within the VVJJ analysis, the tagger consists of consecutive p_T dependent cuts on a number of variables (traditionally mass and D_2) that maximise the significance for selecting boosted vector bosons and rejecting background jets.

⁽¹⁾The resolution is defined as $R^r = [Q_{84}(R^r) - Q_{16}(R^r)] / [2 \times Q_{50}(R^r)]$ and $R^d = 1/2[Q_{75}(R^d) - Q_{25}(R^d)]$ for the mass and D_2 , respectively, where Q_x is the $x\%$ quantile boundary, meaning that Q_{50} is the median. The mass response is defined as $R^r = m_{\text{reco}} / m_{\text{truth}}$, while the residual of D_2 is $R^d = D_{2,\text{reco}} - D_{2,\text{truth}}$, where ‘truth’ and ‘reco’ refer to the truth and reconstructed properties of the jets.

5.3.1 Choice of significance metric

The first order of business is to define a metric for such significance ζ :

$$\zeta = \frac{\varepsilon}{\frac{3}{2} + \sqrt{B}} \quad (5.1)$$

where ε and B are, respectively, the signal efficiency and the number of background events after the cuts are applied. The number 3 is the number of standard deviations corresponding to a one-sided Gaussian distribution⁽²⁾. This definition has numerous advantages compared to traditional significance metrics such as $\zeta = S/\sqrt{B}$ and $\zeta = S/\sqrt{S+B}$. Indeed, using the efficiency ε instead of the number of signal events S that remain after the tagger is applied means that this definition of ζ doesn't depend on a specific signal cross-section but is valid for all signals with the same experimental signature. On top of that, this metric is robust at small values of B unlike the significance definition $\zeta = S/\sqrt{B}$ ⁽³⁾.

5.3.2 Choice of substructure variables and inputs to jet reconstruction

The cuts can, hypothetically, be applied over any number of variables. However, in a traditional optimisation (where every possible combination of variables is examined), the complexity and computational time scale exponentially with the number of variables to cut on. In the previous run of the VVJJ analysis, two variables were used with very good results: the jet mass and D_2 . These variables, described in detail in section 4.6, have been found to provide the highest level of discrimination between signal and QCD if only two variables were to be used [1]). The radiation of a hard gluon can, however, allow background jets to mimic a two-pronged structure very similar to that of vector boson jets. To further improve the sensitivity of the analysis and discriminate against such gluon-initiated jets the original two-variable tagger is re-optimised and extended by adding a third variable to cut on.

Applying simultaneous restrictions on highly correlated variables can lead to diminishing return in the tagging efficiency [94], as such, the third variable should, ideally, be uncorrelated with both mass and D_2 and provide some discrimination between signal and background. In practical terms we see that these requirements can be summarised as looking for a variable that is uncorrelated with the vector boson D_2 and anti-correlated with QCD jets D_2 . The correlation tables shown at the end of section 4.6 hint that two promising variables are KtDR and Angularity. Previous studies [1] also showed that some discrimination could be attained by selecting on the charged hadron multiplicity. This multiplicity can be expressed in the form of the track multiplicity (N_{trk}) of the untrimmed $R = 1.0$ jet, considering tracks with $p_T > 0.5$ GeV consistent with coming from the primary vertex [32]. With the same idea in mind the jet constituent multiplicity N_{TCC} is also tested as a potential variable.

⁽²⁾To be noted: the result of the optimisation is not very sensitive to the exact number of standard deviations chosen [32].

⁽³⁾This becomes clear with a practical example. Let's look at two scenarios: in one we have $B_1 = 10^{-5}$ events and $S_1 = 1$ events, in the other $B_2 = 1$ event and $S_2 = 10$ events. The second scenario should be favoured but with $\zeta = S/\sqrt{B}$ the first one results in a higher significance.

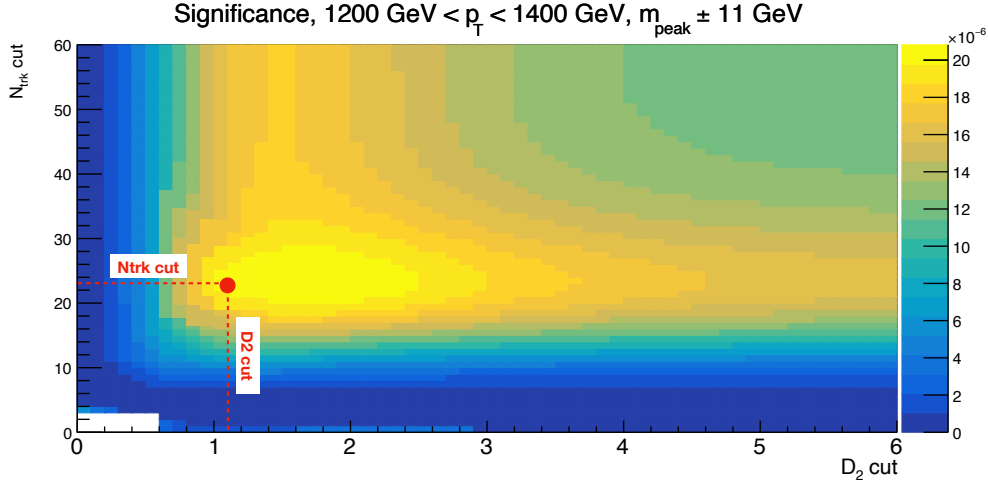


Figure 5.2: Significance for W jets with respect to QCD background jets for a jet with $1.2 < p_T < 1.4$ [TeV]. In red the maximum significance point is shown corresponding to a mass window cut of ± 11 GeV, a D_2 cut of 1.1 and a cut on N_{trk} of 23.

For each transverse momentum bin, the optimal cut values for the three variables are determined by examining all possible combinations⁽⁴⁾ and choosing the one resulting in the highest significance. An example distribution of the significance for TCC W-jets in the truth jet p_T bin [1200, 1400) GeV is shown in Figure 5.2. Since the significance as a function of all three variables cannot be easily visualised due to the three-dimensionality of the representation, the significance is shown for the optimal mass window cut (11 GeV) and as a function of D_2 and N_{trk} cuts. The overall point of maximum significance can be seen in red.

Each cut is applied consecutively meaning that the distribution of D_2 is influenced by the cuts on the mass and the distribution of N_{trk} is influenced by the cuts on the mass and D_2 . The aforementioned distributions for the optimised tagger can be seen in figure 5.3.

To make a complete comparison, nine taggers were optimised: five taggers for TCC jets (four three-variable taggers - Angularity, KtDR, N_{trk} and N_{TCC} - and a standard two-variable tagger) and four taggers for topo-cluster jets (three three-variable taggers - Angularity, KtDR, N_{trk} - and a standard two-variable tagger). The analysis focuses on a momentum range between 1 and 2 TeV. The significance gain for each added variable with respect to the standard two-variable tagger is shown in figure 5.4. For both topo-cluster jets and TCC jets the choice of N_{trk} as a third variable gives the largest gain in significance.

Finally a comparison of the best performing TCC and topo-cluster three-variable taggers and their respective reference two-variable taggers are shown in figure 5.5. The tagger optimised for TCC jets

⁽⁴⁾The simultaneous treatment of the three variables properly accounts for the correlations amongst them.

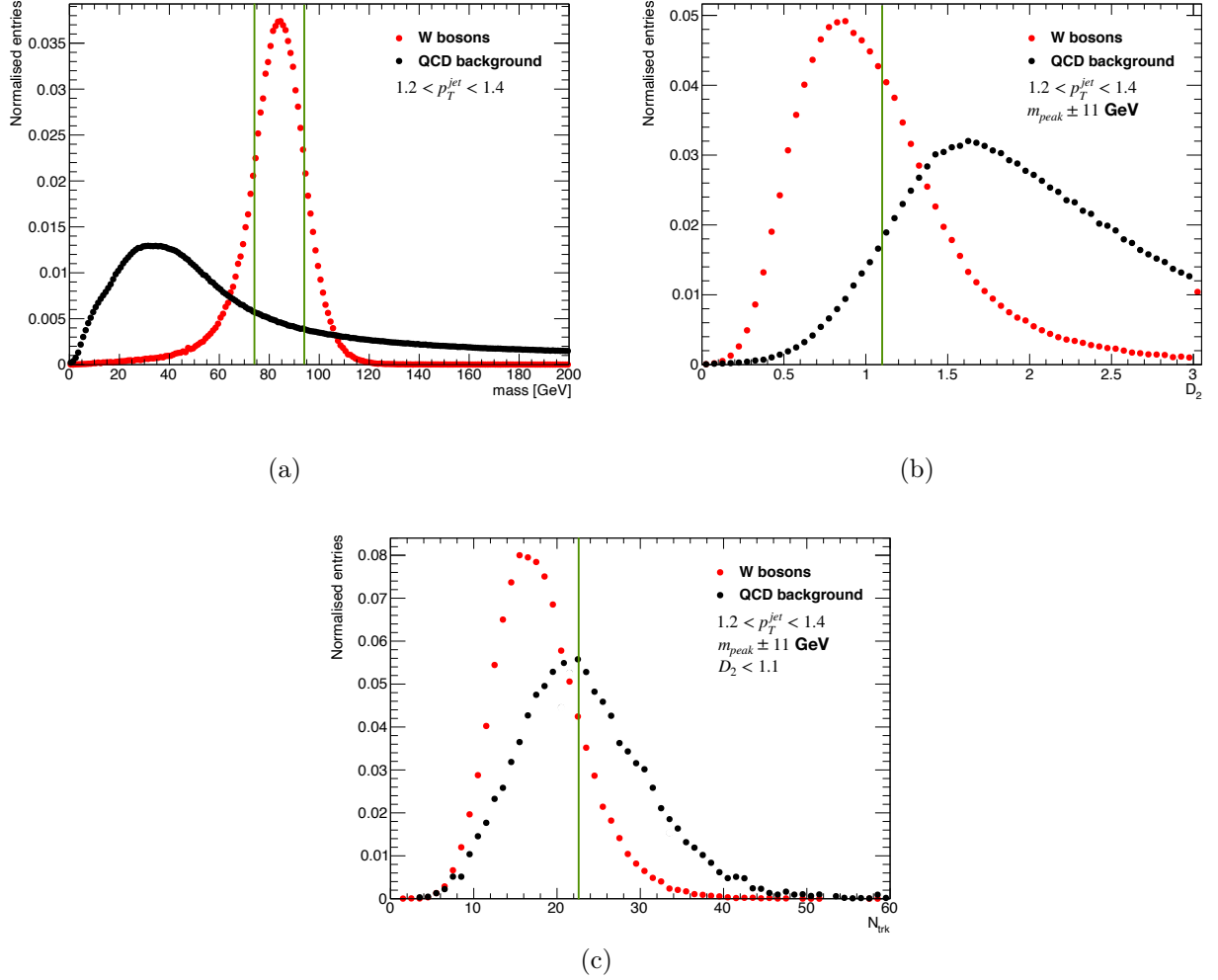


Figure 5.3: In (a) the jet mass distribution. In (b) the D_2 distribution after applying the two-sided cut on the mass. Finally in (c), the N_{trk} distribution after applying the two-sided cut on the mass and the one-sided cut on D_2 . The values are those of jets with $1.2 < p_T^{jet} < 1.4$ TeV. The lines in green indicate the cut values for each variable.

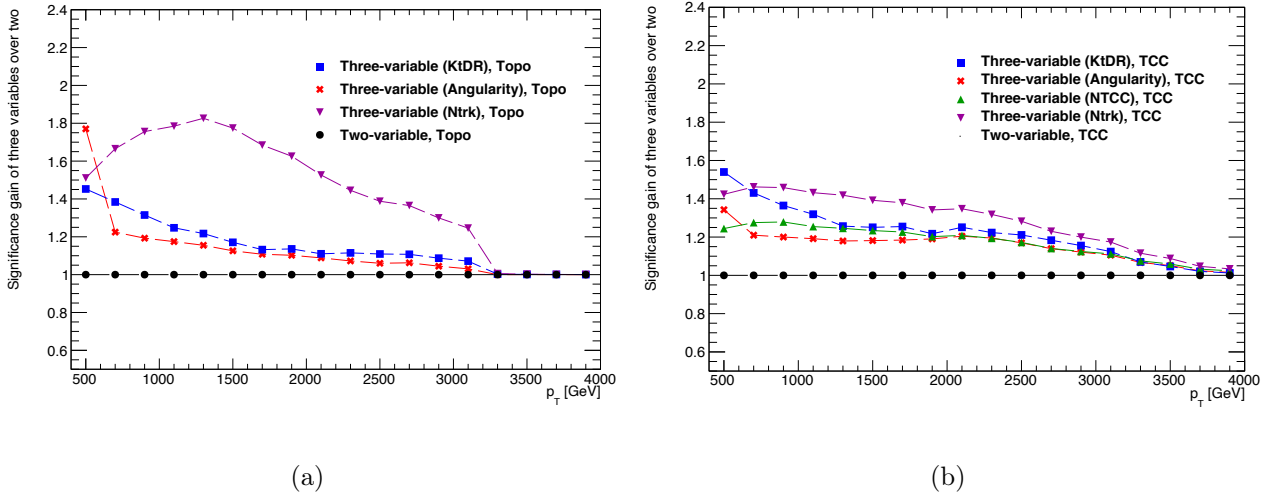


Figure 5.4: In (a), the significance (ζ) gain for topo-cluster jets, defined as $\zeta_{3var}/\zeta_{2var}$, for different choices of the third-variable. In (b), the significance (ζ) gain for TCC jets, defined as $\zeta_{3var}/\zeta_{2var}$, for different choices of the third-variable. The N_{TCC} variable was tested solely on TCC jets as the variable is not defined for topo-cluster jets. The number of tracks outperforms all other variables for both types of jets and especially in the transverse momentum range that is the focus of this analysis ($p_T \approx 1.5$ TeV).

applying cuts on mass, D_2 and N_{trk} shows a striking improvement with respect to all other taggers. The significance gain obtained by using TCC jets instead of topo-cluster jets is also evident as the TCC taggers systematically outperform their topo-cluster counterparts.

5.3.3 Performance evaluation and parametrisation of optimised tagger

Once the optimisation procedure is done and the optimal cut values for each p_T bin are found, parametric functions are used to fit the p_T dependence of the jet mass window, the jet D_2 cut and the jet N_{trk} cut. Indeed, using smooth functions of p_T allows us to avoid bin-edge effects that would result from the use of a discrete number of cut values over the entire transverse momentum range.

For the mass window the following parametrisation is used:

$$F_{mass} = \sqrt{\sigma_E^2 + \sigma_A^2} = \sqrt{\frac{A^2}{(p_T - B)^2} + C^2(p_T - D)^2} \quad (5.2)$$

Where the two sigmas are inspired by the energy and angular resolution respectively. This specific parametrisation in the form of a sum of two resolution terms is physically motivated. For σ_E the function is built to account for the energy resolution contribution in the mass resolution at low p_T . As the transverse momentum increases, the jet energy resolution decreases and so does its contribution to the mass resolution. To first order, the jet energy resolution falls like $1/p_T$, two free parameters (A and B) are added to account for all other effects. At high p_T , on the other hand, the main contributing factor to the mass resolution is the angular resolution σ_A which worsens (increases) due

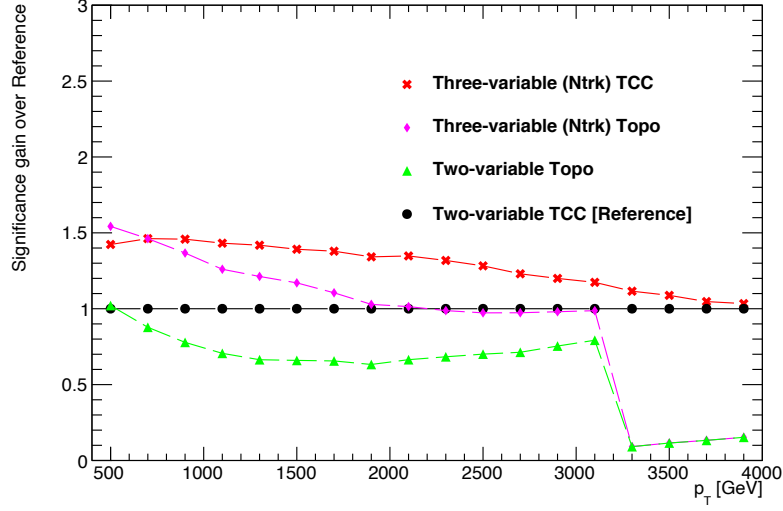


Figure 5.5: Significance gain comparison for various taggers with respect to the reference tagger - TCC two-variable tagger. Both topo-cluster taggers fail at very high p_T (out of the transverse momentum range that is of interest) due to the poor resolution both in mass and D_2 . The TCC three-variable tagger using N_{trk} significantly outperforms the other shown over the entire p_T range of interest.

to the decay products being so collimated that the calorimeter granularity isn't fine enough to resolve the substructure of the event [95].

The D_2 and N_{trk} parametrisations do not have a physical motivation but are modelled as a third degree polynomial⁽⁵⁾ and an exponential function⁽⁶⁾ respectively. The final parametric fitted functions are shown in figure 5.6.

Since the tagger is optimised for maximum significance and doesn't require a fixed signal efficiency nor a fixed background rejection, both of these quantities vary smoothly to maximise the analysis sensitivity [32]. The signal efficiency and background rejection values for the W and Z taggers are shown as a function of p_T in figure 5.7. With the tagger selections applied, we observe a 20% signal efficiency at low p_T which steadily increases with increasing transverse momentum, reaching up to 60%. This behaviour is due to the shape of the multijet background which falls with increasing p_T : with less background contamination the optimal tagger becomes looser to preserve sufficient signal statistics. A complementary behaviour is observed for the background rejection.

5.3.4 Additional event selections applied in the analysis

Before any tagging is applied, events are required to have at least two reconstructed large- R jets of radius parameter $R = 1.0$ originating from the primary vertex. These jets must have $|\eta| < 2.0$, single

⁽⁵⁾ $F_{D_2} = A + Bx + Cx^2 + Dx^3$

⁽⁶⁾ $F_{N_{trk}} = A + e^{Bx+C}$

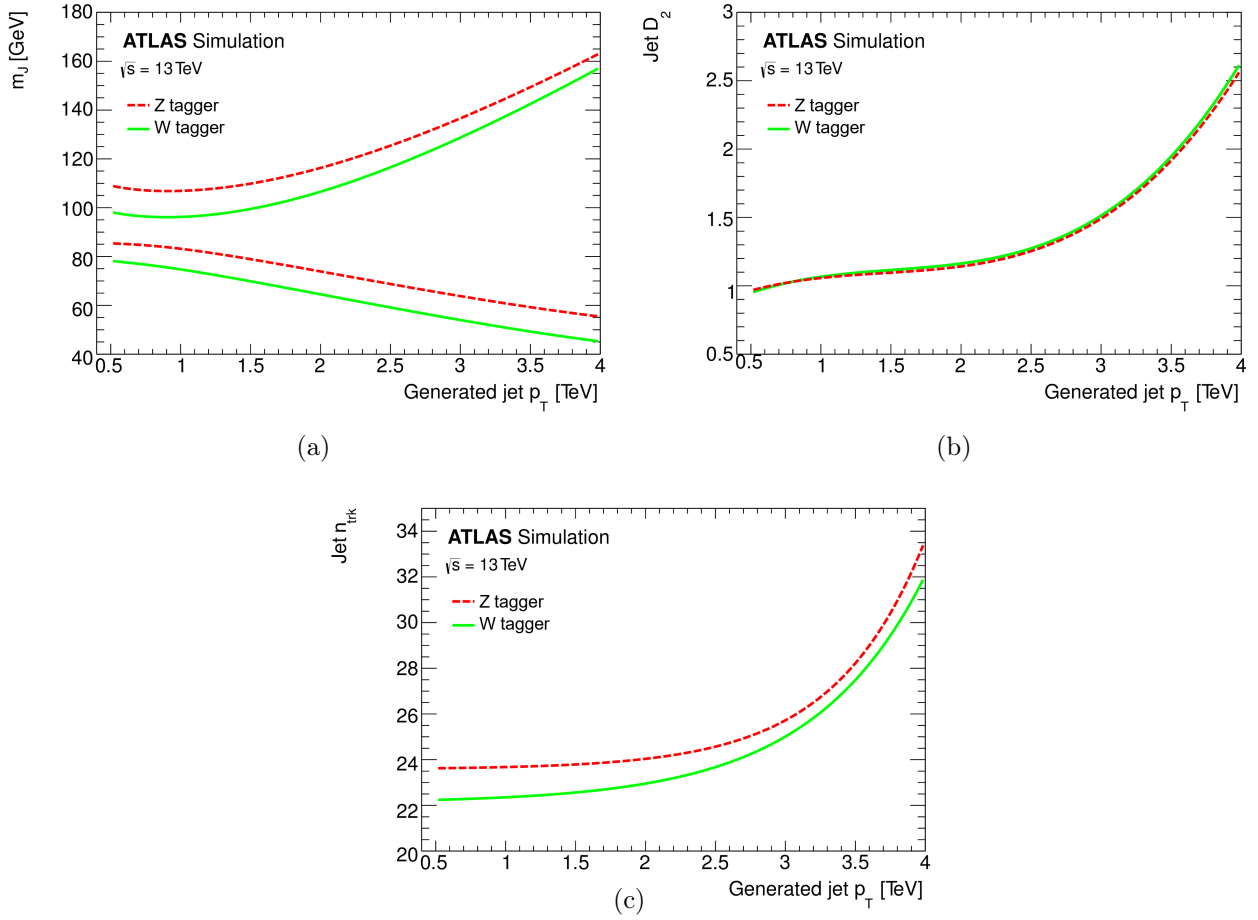


Figure 5.6: In (a) the jet mass window fit. In (b) the D_2 selection fit. Finally in (c), the N_{trk} selection fit. The W and Z tagger are represented in red and green respectively. The tagger is only valid for jets with a transverse momentum between 0.5 TeV and 4.0 TeV and with $|\eta_{jet}| < 2.0$ [32].

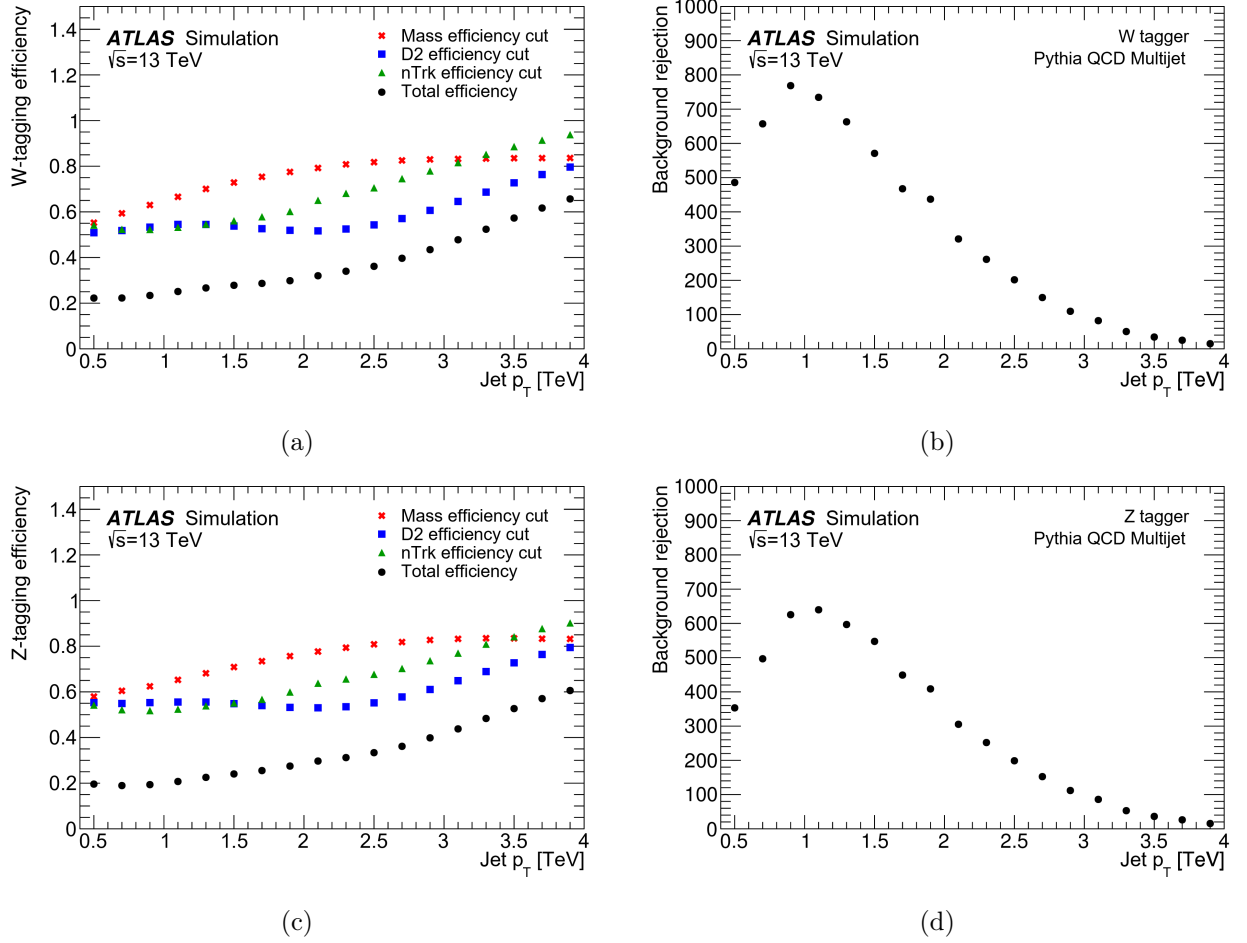


Figure 5.7: The (a) per-boson signal efficiency for the jet mass, D_2 , and N_{trk} selections, as well as the combined efficiency and (b) background rejection ($1/\text{efficiency}$) of the W tagger for HVT $W' \rightarrow WZ \rightarrow qqqq$ and MC simulated multijets as a function of the jet p_T . Corresponding values for the Z tagger are shown in (c) and (d) [32].

jet masses larger than 50 GeV and a dijet invariant mass $m_{JJ} > 1.3$ TeV. The leading jet is required to have a transverse momentum larger than 500 GeV whereas the p_T of the sub-leading one has to be larger than 200 GeV. The leading and sub-leading jets must have a separation in rapidity, Δy_{12} smaller than 1.2 to reduce the multijet contamination. Indeed, this background contribution is mainly produced in t-channel processes with large pseudo-rapidity differences⁽⁷⁾. Finally, a condition is put on the p_T asymmetry of the two jets to remove badly reconstructed jets: $A = (p_{T1} - p_{T2}) / (p_{T1} + p_{T2}) < 0.15$. On top of these selections:

- Events containing a small-R (R=0.4) anti- k_T jet with a momentum larger than 20 GeV and that do not meet the criteria for consistency with pp -collision production are rejected⁽⁸⁾.
- Events with one or more leptons satisfying the following requirements are rejected. Specifically: electrons with $p_T > 25$ GeV and that satisfy the “medium” identification criterion and the “loose” track-based isolation [96], and muons with $p_T > 25$ GeV, $|\eta| < 2.5$ and that satisfy the “loose” track-based selection.

5.4 Boson tagging efficiency

In the previous section we described in detail how to derive the optimal cuts to enhance the separation between signal and background. However, the distributions of the signal efficiency and the background rejection as a function of p_T only give us information on the performance of the tagger over simulated samples. To measure the actual boson tagging efficiency we look at the tagger behaviour in a real data sample enriched in final states with a vector boson plus jets. To obtain said dataset several requirements are imposed:

- Events are required to have two large-R jets with $|\eta| < 2.0$.
- The leading jet transverse momentum must satisfy $p_T > 600$ GeV. This selection criteria allows the sample to be richer in jets with a transverse momentum closer to the values probed in the search.
- Events where a lepton is identified are rejected.

Once the sample is created the D_2 and N_{trk} cuts of the tagger are applied: both jets are tested individually for the presence of a vector boson. To guarantee independence from the signal region, one of the jets is required to pass the tagger selection (i.e. be tagged as either a W or Z boson), and the other is required to fail the cuts.

The mass distribution of the jets passing these selections is the sum of two contributions: that of jets coming from events containing a vector boson and jet, and that of jets that belong to the multijet background. As such the distribution is fitted with a function containing background and signal components, which allows us to extract the rate of V plus jet events by evaluating the integral

⁽⁷⁾Signal events should be produced in s-channel processes with small Δy_{12}

⁽⁸⁾This is done to reduce contamination from calorimeter noise, cosmic rays, beam-induced background, etc...

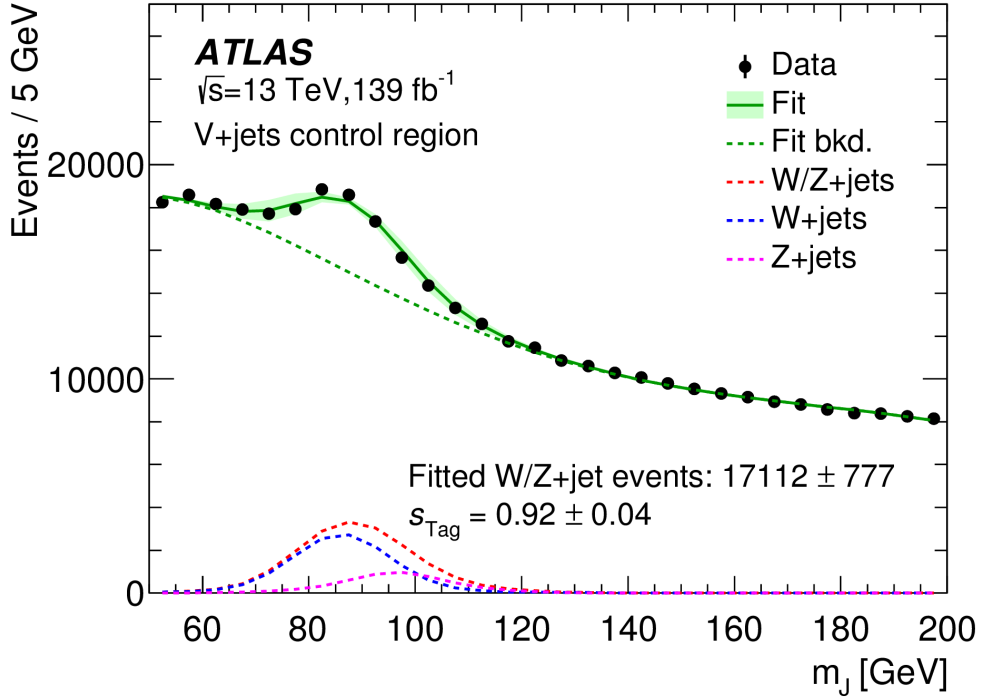


Figure 5.8: Jet mass distribution for data in the region enhanced in $V + \text{jets}$ events after boson tagging based solely on the D_2 and N_{trk} variables. The fit obtained using a background-plus-signal function is also shown in green. The fit uncertainty reflects the uncertainty in shape and positions of the W and Z peaks. At the bottom, the fitted contribution to the observed jet mass spectra from the $V + \text{jets}$ signal is shown. The fitted relative efficiency of the D_2 and N_{trk} selections is $s_{Tag} = 0.92 \pm 0.04$, where the uncertainty is purely statistical [32].

of the fitted signal contribution. The rate of W plus jet events with respect to that of Z plus jet events is fixed from Monte Carlo simulation. The shape of the fit function is the sum of the parametrised background contribution (fit to data using a fourth-order exponentiated polynomial) and the modelled $V + \text{jet}$ contribution (represented by a double gaussian - one for W and one for Z - distribution whose shape parameters are determined from simulated data). The fitted mass distribution can be seen in figure 5.8.

The overall efficiency of the N_{trk} and D_2 cuts in data (ε_{data}) compared with Monte Carlo simulation (ε_{MC}) is measured to be $s_{tag} = \varepsilon_{data}/\varepsilon_{MC} = 0.92 \pm 0.04$ where the uncertainty is purely statistical. Several uncertainties need to be added to the statistical one.

First the so called MC closure of the fit has to be probed: signals of various strength are injected in simulation, then the background-plus-signal fit function defined earlier is applied and the ability of our parametrisation to correctly extract the overall $V + \text{jet}$ contribution (or yield) is tested. Since good agreement is found (uncertainty of 2%) the method is deemed reliable. Comparisons between simulated samples and data allow us to test other possible sources of contributions to the overall yield measured in V plus jets. In particular a contribution from $t\bar{t}$ background of roughly 5% is

subtracted. A theoretical uncertainty of 10% in the modelling of the V +jets cross section for a V with a transverse momentum $p_T \approx 600$ GeV is also accounted for [97]. Uncertainties for MC closure, $t\bar{t}$ contributions, theory and fit are calculated. Finally, since the efficiency is extracted for vector bosons with transverse momentum starting at 600 GeV but the analysis search p_T range reaches 3.5 TeV, the dependence of the modelling on the jet transverse momentum needs to be evaluated. This is done by comparing the distributions of D_2 and N_{trk} as a function of p_T between data and Monte-Carlo simulations. The residual mismodelling is accounted for by adding an 5% uncertainty to s_{tag} [32].

The final value of the relative efficiency of the N_{trk} and D_2 cuts is therefore:

$$s_{Tag} = 0.92 \pm 0.04(\text{stat}) \pm 0.02(\text{closure}) \pm 0.03(\text{tt}) \pm 0.02(\text{fit}) \pm 0.05(\text{pT range}) \pm 0.10(\text{theory}) \quad (5.3)$$

This relative efficiency is used to scale MonteCarlo simulated signal events and its uncertainty represents the uncertainty of the W/Z tagger in simulation.

It is interesting to note that the polarisation in the V plus jet sample is primarily transverse which is in contrast with some of the signal models probed with the search. However since no other physics process allows us to probe the differences between data and simulation at this energy scale it is assumed that the modelling of polarisation in MC samples is good enough that the scale-factor extracted can be applied generally.

5.5 Event selection efficiency

Events that pass the tagger selection are divided in five non-exclusive signal regions (SRs). Three regions are used to probe the individual decay channels and contain events with two jets tagged as WW , ZZ , or WZ . The remaining two regions combine the first three in order to have a higher sensitivity to the signal described in the theoretical scenarios of section 5.1, and require the jets in the events to be identified as either WZ or WW for the first region and WW or ZZ for the second region. The only difference between the regions are the boson-tagging requirements whereas all the other pre-selections are the same. Generally the jet with the highest mass is considered as a candidate for the boson with the highest mass⁽⁹⁾.

The event selection efficiency is defined as the ratio between the number of selected events and the total number of generated events and is shown in figure 5.9 as a function of the resonance mass. From the image shown, boson tagging is the step that more drastically reduces the amount of signal selected. It is, however, also the step that provides the largest suppression of QCD background, as can be seen in figure 5.10.

⁽⁹⁾For instance, for the WZ selection, the leading jet mass and the sub-leading jet mass must fulfill the Z and W tagger requirements respectively.

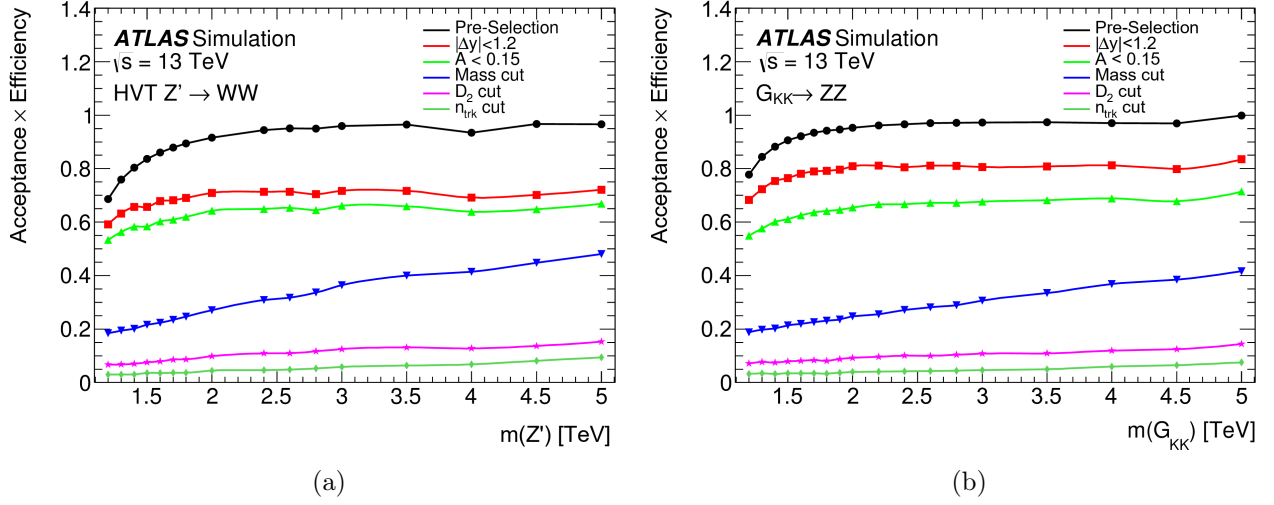


Figure 5.9: In (a), the event selection efficiency as a function of mass at different stages of the selection procedure for HVT $Z' \rightarrow WW$. In (b), the event selection efficiency as a function of mass at different stages of the selection procedure for $G_{KK} \rightarrow ZZ$. The selections are applied in sequence and include all the selections described over the course of this chapter [32].

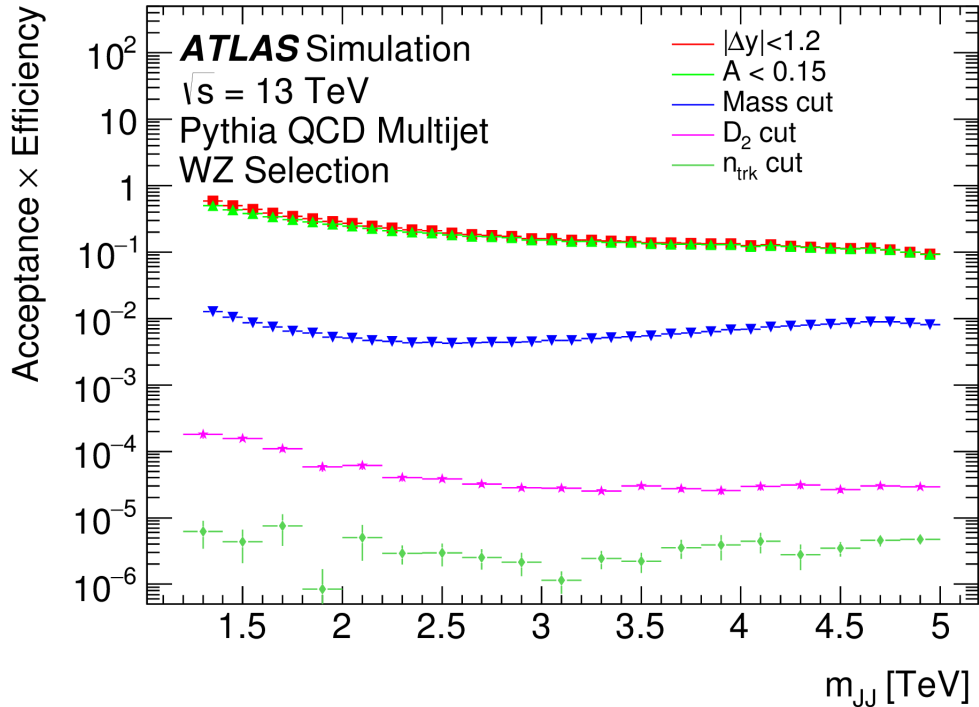


Figure 5.10: The efficiency for the selection of the MC multijet background as a function of dijet mass at different stages of the selection procedure. The selections are applied in sequence and include all the selections described over the course of this chapter [32].

5.6 Background parametrisation

The VVJJ analysis looks for a localised excess over a smoothly falling dijet mass distribution as expected by the Standard Model. The correct estimation of the background contribution is therefore of key importance. For the VVJJ analysis the background is estimated empirically by performing a binned maximum-likelihood fit of the following parametrised form to the dijet mass spectrum:

$$\frac{dn}{dx} = p_1(1-x)^{p_2-\xi p_3} x^{-p_3} \quad (5.4)$$

where $x = m_{JJ}/\sqrt{s}$, p_1 is a normalisation factor, p_2 and p_3 are dimensionless shape parameters, and ξ is a constant [32]. The validity of the fit parametric function is tested in a dedicated region in data using an ABCD-like method. To do so, four orthogonal (exclusive) regions are defined: region A contains events in which both jets are boson-tagged and the separation in rapidity $|\Delta y_{12}|$ is larger than 1.2, region B has the same tagging requirements but $|\Delta y_{12}| < 1.2$, region C requires the jets to not be boson-tagged and $|\Delta y_{12}| > 1.2$ and region D has both jets not boson-tagged and $|\Delta y_{12}| < 1.2$. The two regions with a large rapidity separation (A and C) are used to derive per jet weights to be applied to go from region D to region B which corresponds to the nominal signal region. The idea is to test the validity of the test function on a control region that highly resembles that of the signal region thus obtaining robust results without biasing the analysis.

The studies performed confirm the ability of the chosen background fit function to describe the expected dijet mass spectra as can be seen in figure 5.11[95].

5.7 Systematic uncertainties

The uncertainties affecting the background modelling are taken directly from the errors in the fit parameters of the background estimation procedure described in the previous section. The systematic uncertainties due to detector effects and Monte Carlo mismodelling are expressed in terms of nuisance parameters. The main sources of uncertainty in the signal modelling arise from uncertainties in the large-R jet tagging efficiency and the jet p_T calibration [32].

In particular, the uncertainty in the jet transverse momentum calibration is especially significant since it can shift the peak of the searched for resonance and would therefore affect the significance of an observed excess [95]. It is estimated by using the R_{trk} method (see section 4.5.3), that is by computing the track to calorimeter double ratios between data and simulation [99]. Since the two ratios are expected to be the same, any observed difference is interpreted as an uncertainty. This method however, relies on the fact that the two ratios should be uncorrelated. This is no longer true for TCC jets since they combine tracker and calorimeter information. To correct for this, an upper limit to the correlation is estimated by comparing the double ratios obtained using TCC jets and topo-cluster jets and is found to be at the percent level, which is taken as an uncertainty. Additional uncertainties due to the track reconstruction efficiency, track impact parameter resolution and track fake rate are taken into account. The size of the total jet p_T scale uncertainty varies with jet p_T , as

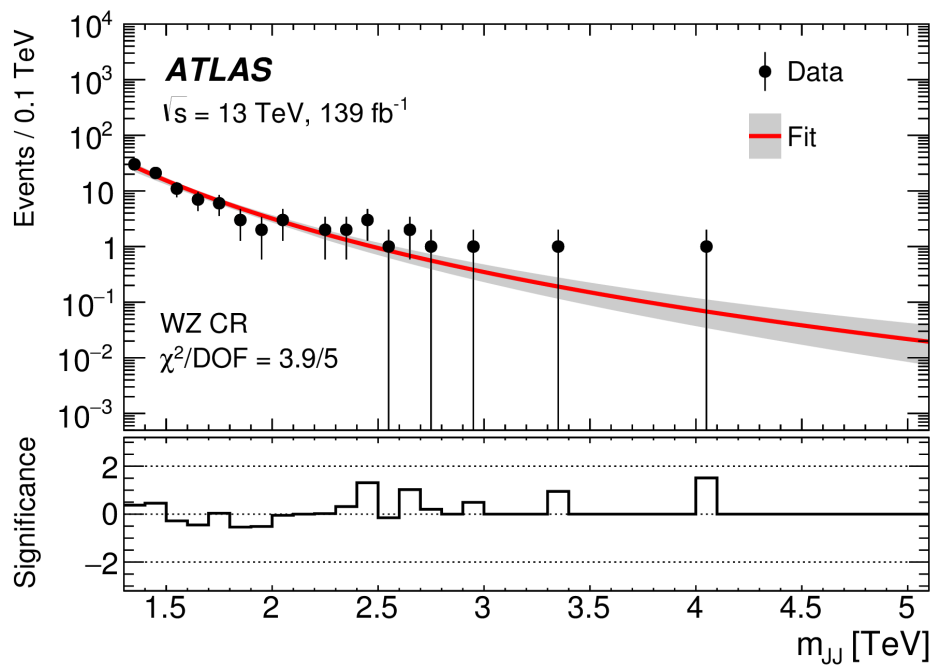


Figure 5.11: Comparison between fitted background shape and the m_{JJ} spectra in an example WZ fit control region in data [32]. The fitted background distribution is normalised to the data shown in the displayed mass range. The shaded bands represent the uncertainty in the background expectation calculated from the maximum-likelihood function. The lower panel shows the significance as defined in Ref. [98].

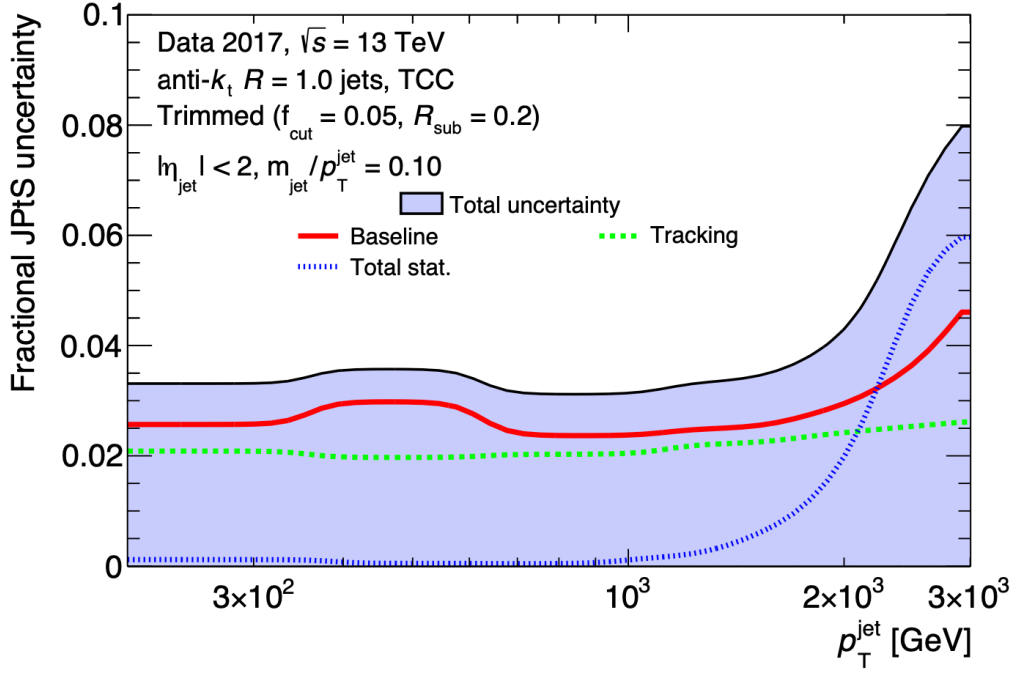


Figure 5.12: Fractional jet p_T scale systematic uncertainty components as a function of p_T for jets with $\eta < 2$ and $m_{jet}/p_T^{jet} = 0.10$ [100]. The final uncertainty values reported in the text translate this plot to all mass ranges.

can be seen in figure 5.12, and is between 2.5% and 5% for the full range [32].

Uncertainties in the measurement of the jet p_T resolution would cause errors in the measurement of the width of any observed resonance and affect the signal selection efficiency. It is evaluated event-by-event by rerunning the analysis with an additional Gaussian smearing to the jet's input p_T . The width of the Gaussian is an absolute 2% per jet and is symmetrised [32].

The boson tagging efficiency is affected by the uncertainty in the jet mass scale and resolution as it influences the observed jet mass. Any uncertainty in the value of the boson-tagging discriminant $D2$ or N_{Trk} , would also affect the selection efficiency of the analysis. A scale-factor for the W/Z-tagging efficiency is derived as described in section 5.4 and the uncertainty in the scale-factor is assigned as a two-sided variation in the yield [32].

Finally the uncertainty on the integrated luminosity between 2015-2018 is 1.7% [101] and the uncertainty from the trigger efficiency is found to be negligible⁽¹⁰⁾.

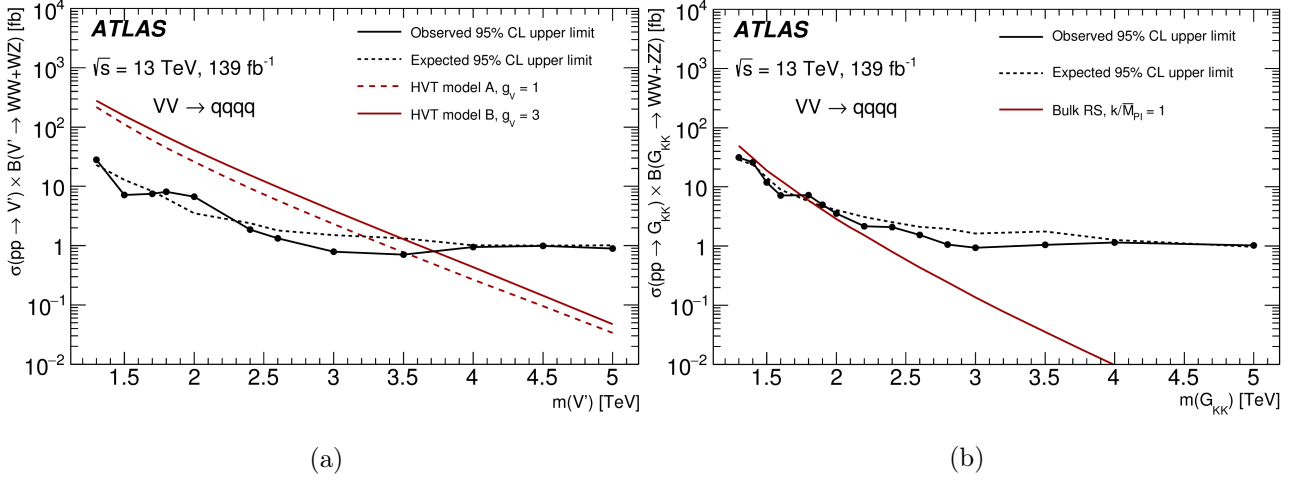
⁽¹⁰⁾The minimum requirement on the dijet invariant mass of 1.3 TeV guarantees that the trigger is fully efficient.

5.8 Results

No events are observed above 5 TeV. In the combined $WW + ZZ$ and $WZ + WW$ regions 113 and 119 events are observed respectively. Due to the non-exclusive selections of the boson taggers, about half the events satisfying the WW selection also satisfy the ZZ selection. The highest mass event is at 4.4 TeV and is the same for both signal regions, it is compatible with the background expectation in the high mass region [32].

The significance of observed excesses over the background-only prediction is quantified using the local p_0 -value, defined as the probability of the background-only model to produce a signal-like fluctuation at least as large as that observed in the data. The most extreme p_0 has a local significance of 1.8 standard deviations, and is found when testing the HVT $W' \rightarrow WW$ hypothesis at a resonance mass of 1.8 TeV. This is within the expected fluctuation of the background [32].

In the absence of a significant excess, 95% confidence level (CL) upper limits on the production cross-section times the branching fraction to diboson final states are set for the signals expected from the benchmark scenarios described in section 5.1. In this case, all the systematic uncertainties described in section 5.7 are taken into account. It is interesting to note that the uncertainty on the efficiency of the W/Z taggers is dominant at lower masses whereas the limiting factor at high masses is the uncertainty on the background parametrisation. The cross-section limits estimated for the different benchmark signal models in the two signal regions with the largest sensitivity (that is the $WW + ZZ$ and $WZ + WW$ signal regions) can be seen in figure 5.13.



| Model | Signal Region | Excluded mass range [TeV] |
|------------------------------------|---------------|---------------------------|
| Radion | WW | 1.3–2.8 |
| | ZZ | 1.6–2.1 |
| | $WW + ZZ$ | 1.3–3.0 |
| HVT model A, $g_V = 1$ | WW | 1.3–2.9 |
| | WZ | 1.3–3.4 |
| | $WW + WZ$ | 1.3–3.5 |
| HVT model B, $g_V = 3$ | WW | 1.3–3.1 |
| | WZ | 1.3–3.6 |
| | $WW + WZ$ | 1.3–3.8 |
| Bulk RS, $k/\overline{M}_{Pl} = 1$ | WW | 1.3–1.6 |
| | ZZ | none |
| | $WW + ZZ$ | 1.3–1.8 |

(c)

Figure 5.13: Observed and expected limits at 95% CL on the cross-section times branching ratio for $WW + WZ$ production as function of $m_{V'}$ in (a), and for $WW + ZZ$ production as a function of the Bulk RS graviton m_G in (b). The predicted cross-section times branching ratio is shown as dashed and solid lines for the HVT models A and B respectively in (a), and as a solid line for the bulk RS model in (b). Finally in (c) the observed excluded resonance masses (at 95% CL) in the individual signal regions for the HVT, bulk RS and radion models [32].

6

Machine Learning based classification

Contents

| | | |
|------------|---|------------|
| 6.1 | Machine Learning Basics | 107 |
| 6.1.1 | Datasets | 108 |
| 6.1.2 | Inputs to machine learning algorithms | 109 |
| 6.2 | Multi-Layer Perceptrons and Deep Neural Networks | 111 |
| 6.2.1 | Perceptron | 111 |
| 6.2.2 | Deep Neural Networks | 113 |
| 6.2.3 | Activation functions | 114 |
| 6.2.4 | Convolutional Neural Network | 115 |
| 6.2.5 | Hyperparameters and Regularisation | 117 |
| 6.3 | Edge Convolution | 118 |
| 6.4 | Software | 120 |

Machine learning can be shortly described as the field concerned with automatically creating a model of a phenomenon from a certain amount of examples of said phenomenon [102, 103]. This very general description hints at the variety of uses machine learning algorithms can have. In particle physics one of the main application is the use of ML algorithms to correctly identify the nature of the particles produced in the collisions at the interaction point. Indeed, as was discussed in the previous chapter, a full scan over all possible combinations of cuts on variables can only be performed on a small number of variables. Machine learning algorithms, and, in particular, deep learning algorithm, can tackle the level of complexity that arises from expanding the analysis to a larger number of variables. In this chapter we will discuss the various concepts necessary to understand how a Machine Learning based tagger is developed.

6.1 Machine Learning Basics

The types of learning algorithms are separated based on what information the algorithm has access to. The problem we are trying to solve, that is, the identification of a jet's origin particle, falls under the category of supervised learning for a classification task.

Supervised classification consists of a supervised learning algorithm tasked with automatically assigning a label to an unlabeled example. In supervised learning, each of the training examples consists of a set of features that provide some type of information about the object that needs to be identified, and the label of said object. For classification tasks, the label denotes the class the object belongs to. For instance, in the case of boson tagging, the features could be the jet substructure variables and the classes could be 1 for jets originating from vector bosons (W and Z confounded) and 0 for jets originating from any other particle. The example just mentioned is an example of binary classification, meaning that only two classes are taken into consideration. If there are three or more classes, as is the case if we want the vector boson jets to be labeled as W , Z or other, we talk about multiclass classification.

In short, a supervised classification algorithm builds a model that, given an unlabeled input, is able to map the input features to the correct output label. If, in doing so, the algorithm learns the parameters (or weights) of the model directly from the features we talk about shallow learning. In deep learning, instead, new, derived features are created from the input features by introducing multiple learning layers between the input features and the output label.

6.1.1 Datasets

To build an algorithm, at least three datasets need to be created: one for **training**, one for **validation** and one for **testing**. To do so the raw data is shuffled and randomly split into the three datasets with different proportions, the training dataset being the largest one. What are each of these datasets used for? The training dataset is the one used to build the model. The examples in the testing and validations datasets CANNOT be accessed while training the algorithm. These last two datasets are used, as their names imply, to **validate** the choice of the model and all its user defined parameters (hyperparameters) and to **test** the performance of the built model before deployment.

The reason behind keeping some examples from being seen by the model while it's being built is the same that motivates the definition of all aspects of a physics analysis before looking at the actual data: bias. We do not want the model to be good at labelling only what it has already seen, rather, we want the model to correctly predict unseen data (or with the example of a physics analysis: we want to observe if a real excess is present, not find a way to extract an artificial excess from data). A model with a large number of parameters can easily “memorise” all the examples seen and therefore learn a model that perfectly adheres to the training data. When applied to unseen data, such a model will be useless. We do not want perfect performance over the training dataset, we want good performance over unseen data. The additional split between validation and test datasets is justified by the fact that the validation dataset is also in some way used in the process of building the model since we use it to choose which is the “best” model. A third completely independent testing dataset allows for further confirmation of the behaviour of the model on general data.

The general concepts described in the previous paragraph are expressed through the use of two terms: **underfitting** and **overfitting**, and can be seen in figure 6.1. Underfitting is the inability of the model to predict with a certain amount of certainty the labels of the data that it is trained on. This can be due to the use of a model that is too simple or to the use of features that do not contain enough

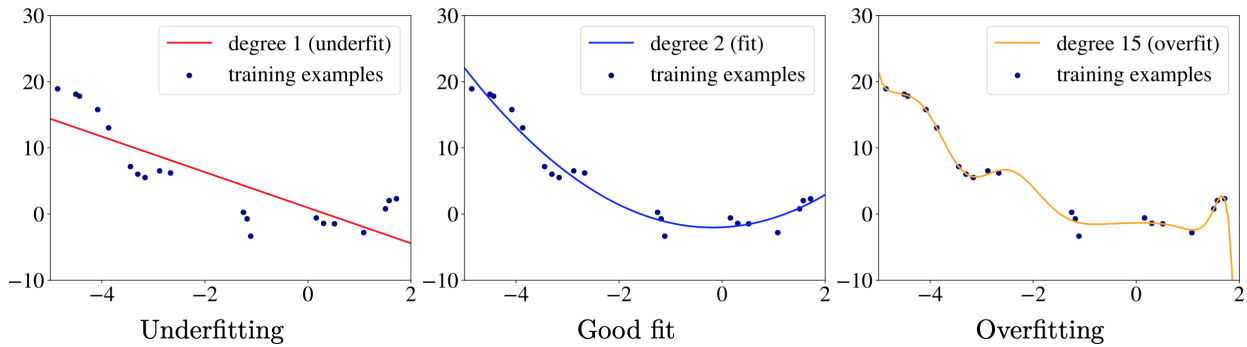


Figure 6.1: Examples of underfitting (linear model), good fit (quadratic model), and overfitting (polynomial of degree 15) [102].

predictive information [102]. Overfitting, on the other hand, is when the model predicts very well the training data but performs poorly on the other two datasets. This happens for a number of reasons but mainly due to using an overly complex model or a high number of features with a small number of examples.

Overfitting can be reduced by finding good hyperparameters and by applying regularisation methods. Very generally, regularisation is a term that encompasses all methods that, in one way or the other, simplify the model, during the training, to avoid overfitting [102]. Hyperparameters, on the other hand, are properties of an algorithm that are not learnt: these properties are user defined. To find which combination of hyperparameters results in the best performance, an optimisation of the possible values can be done by scanning all possible combinations by hand. This procedure is called hyperparameter tuning. The tuning can be more or less fine: most of the gain obtained from changing the parameters is of the order of less than one percent but sometimes changing a few key values can result in quite a dramatic improvement. Since there is no end to all possible combinations, a compromise between time cost and performance gain needs to be found. All hyperparameters used in the algorithms that are the object of study of the rest of this thesis will be described in detail later in this chapter, in section 6.2.

6.1.2 Inputs to machine learning algorithms

Contrary to what one might think, data preparation (the preparation of the “examples” to learn/predict from) is one of the most crucial and time intensive steps in the development of an efficient solution to a classification problem. Data preparation consists in transforming raw data into a set of labeled examples and generally requires domain knowledge. The importance of domain knowledge is clear when working with data from the ATLAS detector. Indeed, given the complexity of the phenomena observed and of the reconstruction procedure, and the sheer amount of information that is gathered, it would be a monumental task for an algorithm to acquire the knowledge necessary to predict the particles observed in a collision from raw detector data. Domain knowledge allows us to reduce the problem complexity by focusing on a smaller part of the identification process and to work with objects

that have been engineered by particle physicists (such as jets, jets constituents and jet moments). There are several steps in data preparation and they differ depending on the algorithm used, for classification problems the following constitutes a rather standard procedure:

- First, the problem needs to be well defined: what are we trying to classify? In our case we limit the classification to large-R jets reconstructed with the anti- k_T algorithm labeled with the jet's origin particle (truth information is available due to the fact that the samples are MC simulated).
- Second, the number of examples per class needs to be balanced within the dataset. If one class is highly represented whereas the other is scarcely represented, any algorithm, in case of doubt, will favour predicting the class containing the largest number of examples as this will result in higher overall accuracy⁽¹⁾. To mitigate unbalanced datasets several strategies can be applied. In particular, if the number of examples is large (as is the case for data collected and simulated at the LHC) the easiest and most efficient solution is to downsample the training and validation datasets so that the number of examples per class is the same.
- Third, the features that we want to use to describe the problem need to be selected. The information contained within a jet is large and can be divided in two categories: low-level information and high-level information. The latter consists of information within which simpler levels of knowledge are already condensed, a good example of such input would be jet substructure variables. On the other hand, an algorithm can be trained over a more fundamental - or low-level - representation of the object under study and then asked to learn more complex concepts from there. This would be the case of using the four-vectors of the constituents of jets as input. It is important to note that redundancy in the information given (for instance using features that are highly correlated) can be detrimental to the performance of the algorithm, due to this, the use of a reduced number of features can often lead to better results in the overall performance.
- Fourth, many algorithms require information in the form of matrices (tensors) of fixed size, meaning that missing values or inputs of variable length need to be dealt with in various ways. For missing input features two main solutions are proposed: if the fraction of examples with missing features is small, the simplest solution is to discard them, otherwise, the undefined values can be substituted by a random plausible value or the average value for said feature over the available examples. For inputs of variable size, like the jet constituents whose number varies between events, a padding can be applied. Imagine that the algorithm expects a matrix of size $(N_{jets}, N_{constituents}, N_{features})$, if the number of constituents for the jet in question, $N_{constituents}^{jet}$, is smaller than $N_{constituents}$, a number N_{ghost} of ghost constituents can be added to the jet where $N_{ghost} = N_{constituents} - N_{constituents}^{jet}$. These ghost constituents have a random distribution in η and ϕ (within the jet) and an infinitesimally small value in p_T and energy. The addition of ghosts doesn't change the properties of the jet and is regularly done in different reconstruction steps (see section 4.2).
- Fifth, many algorithms benefit from standardisation: that is the procedure during which the feature values are rescaled so that their distribution is centred at 0 ($\mu = 0$ where μ is the average

⁽¹⁾Sum of true positives and true negatives divided by the sum of false positives and false negatives.

value of the feature over all examples in the dataset) and the standard deviation is $\sigma = 1$. In short terms, this means changing the values so that, once scaled, their distribution has the properties of a standard normal distribution.[102]. It should be noted that the scaling factors are determined from the training dataset, and they are not re-derived but rather applied identically to the validation and testing datasets.

Now that we have defined the datasets and inputs needed to build a properly functioning learning algorithm, we can delve into the mechanics of how an algorithm actually “learns”. Neural networks will be used to illustrate the learning procedure as they are the main focus of this thesis. More complex architectures, based on neural networks are also described in the next sections.

6.2 Multi-Layer Perceptrons and Deep Neural Networks

Neural networks get their name in that they were built to emulate the way actual neurons work. Our brains must process an incredibly large amount of information that comes from the environment we live in and they do so by using a large interconnected network of neurons. Each neuron cell sums all inputs coming from other neurons and if said sum is larger than a certain threshold, the neuron fires a signal that is sent to other neurons. As such, the first step in emulating such a structure is to define the single artificial neuron or perceptron.

6.2.1 Perceptron

A depiction of an artificial neuron can be seen in figure 6.2. In the brain analogy, the input features vector \vec{x} represent the input coming from N other neurons. Each input x_i has a multiplicative weight w_i (from the weight vector \vec{w}) that represents the relevance of the input information and an additive bias term b_i . The neuron then computes the weighted sum of the inputs ($\sum = \vec{x} \cdot \vec{w} + b$) and feeds it to the activation function f . The activation function represents the “decision” to fire a signal and many functions can be chosen with various pros and cons.

But how does such a structure learn? All machine learning algorithms consist of at least three key elements: a loss function, an optimisation criterion - cost function - which is based on the loss function, and a method to systematically extract the necessary information from the training data and solve the optimisation problem. We call training the procedure during which the parameters (weights and biases) that minimise the cost function are found.

The loss function is a measure of penalty given when the algorithm makes a mistake, that is, when the output of the algorithm doesn’t coincide with the target output. The loss function is calculated for each training example. The goal of a learning algorithm however, is not that of modelling a single example but rather to learn a general model for all inputs of a certain category. To do so, the cost function generalises the loss to the entirety of the training dataset by using a pooling function (such as calculating the average loss over the entire training dataset or its sum). Finally an iterative method is needed to find the parameters that minimise the cost function, the most widely used amongst this type of methods is called **gradient descent**.

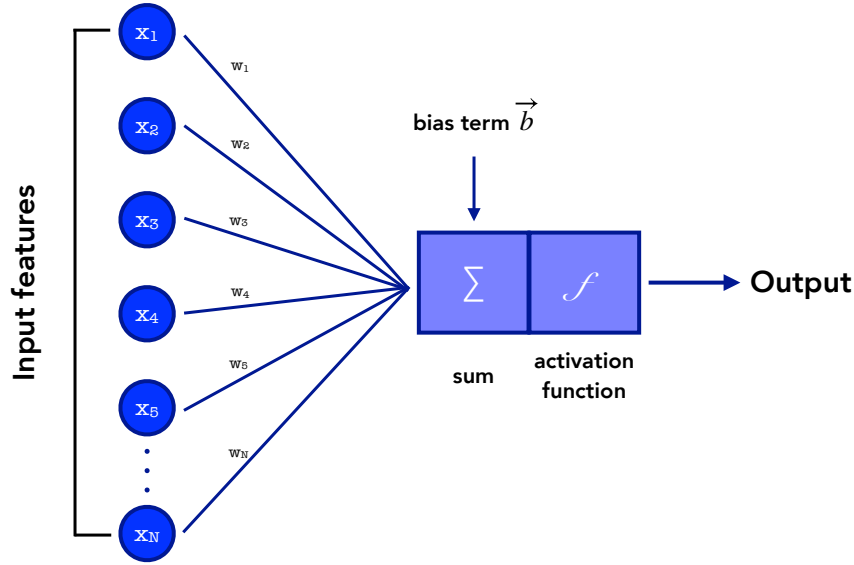


Figure 6.2: Model of a perceptron.

The gradient descent method starts by randomly initialising the algorithm parameters (weights and biases). From there it changes them by an amount that is proportional to the partial derivative of the cost function with respect to each single parameter. By doing so the parameters evolve in the direction that goes towards the minimum of the cost function l . The process is iterative and proceeds in epochs. An epoch consists of one pass over the entire training dataset [104]. Finally, the learning rate α represents the multiplication coefficient in the update:

$$\begin{aligned} w_i &\leftarrow w_i - \alpha \frac{\delta l}{\delta w_i} \\ b_i &\leftarrow b_i - \alpha \frac{\delta l}{\delta b_i} \end{aligned} \quad (6.1)$$

In particular, for a perceptron, the above formula becomes:

$$\begin{aligned} w_i &\leftarrow w_i + \alpha g'(\sum)(T - A)x_i \\ b_i &\leftarrow b_i + \alpha g'(\sum)(T - A)x_i \end{aligned} \quad (6.2)$$

where g' is the derivative of the activation function, α is the learning rate, A is the output given by the perceptron and T is the target output of the perceptron. The perceptron learns by updating each weight w_i and each bias b_i up until when the difference between the perceptron output and the target output is equal to zero. Intuitively, what the perceptron does is to separate the training examples in two categories: those that cause it to fire and those that don't [105]. Since the sum is a linear combination of weighted input features, the algorithm is finding the line that best separates the two

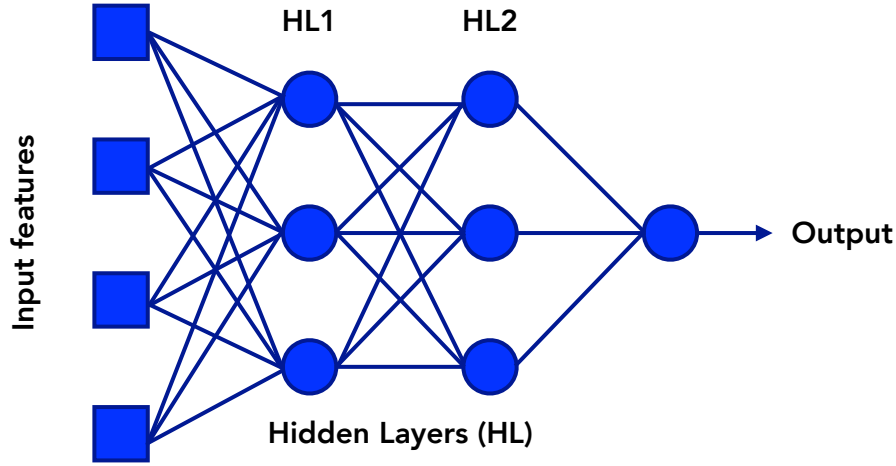


Figure 6.3: Architecture of a deep neural network.

classes. The main limitation of perceptrons is that most problems are not linearly separable. To solve these limitations deep neural networks are introduced.

6.2.2 Deep Neural Networks

Deep Neural Networks (DNN) are neural networks with at least two hidden layers. Nowadays when talking about DNNs, one generally refers to fully-connected DNNs, an example of which can be seen in figure 6.3. In this type of architecture each input is given to every node in the first hidden layer (HL1), in turn, the output of each node in the first hidden layer becomes the input to every node in the second hidden layer (HL2), finally, the output of the second hidden layer are given as inputs to the last node, this last step then amounts to a simple perceptron.

With a multilayer architecture however, the learning algorithm needs not only to learn the parameters between the output layer and the second hidden layer but also the parameters between each hidden layer and between the first hidden layer and the input layer. To solve this problem **back propagation** is used. To tune these “in-between” parameters as shown in equation 6.2 we would need to know the error the network is making at each layer, but the only available error is computed once the output is known. The idea is therefore to use the same equations but to propagate the error found at the output layer, backwards.

In practice, the architecture of the model described above is none other than a nested function f_{NN}

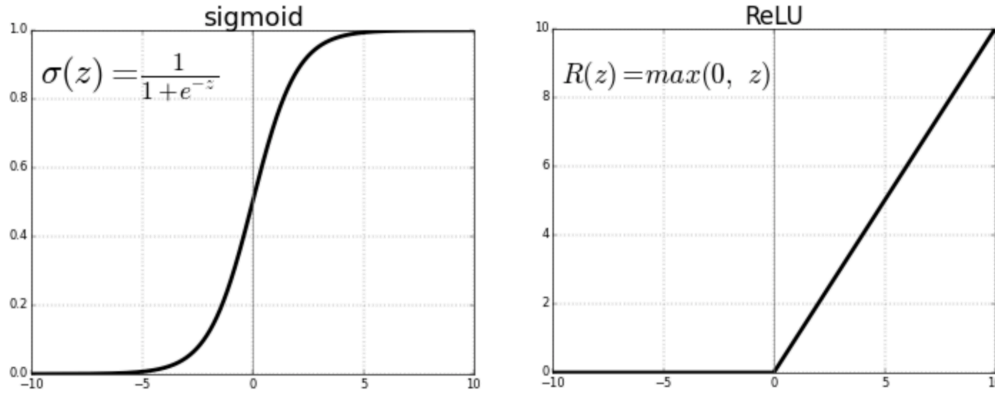


Figure 6.4: On the left the sigmoid activation function and on the right the ReLU function.

mapping inputs \vec{x} to an output y . For the example shown in figure 6.3 we have⁽²⁾:

$$y = f_{NN}(\vec{x}) = f_3(\mathbf{f}_2(\mathbf{f}_1(\vec{x})))$$

$$\mathbf{f}_l(\vec{z}) = \mathbf{g}_l(\mathbf{W}_l \vec{z} + \mathbf{b}_l)$$

Where l is the layer index (1, 2 and 3 in our example), \mathbf{g}_l is the activation function for the layer in question, \mathbf{W}_l is the weight matrix⁽³⁾ and \mathbf{b}_l is the bias term. At the beginning of the training, the weights are randomly initialised using the Glorot uniform initialiser which has been seen to improve the performance of the network. The output is computed as the chain result of the various activations for each layer. With the output known, the error can be back propagated to the nested layers and the process continues until the weights are stable and the maximum performance of the algorithm has been achieved [102, 106].

6.2.3 Activation functions

The choice of the activation functions for the layers of the network is very important: they are the only elements in the model that break the linearity and thus allow the model to extract non linear correlations between the inputs and the output.

The activation functions of the inner layers need to be chosen in order to avoid the problem of vanishing gradient. We refer to vanishing gradient as the problem that arises when the partial derivatives of g are so small that the training stops. This often happens with functions that plateau,

⁽²⁾For simplicity of notation we only use the vector arrow notation for the input vector, matrices are shown as bold capital letters and all other bold letters represent vectors.

⁽³⁾In a multilayer scenario the weights are represented by a matrix instead of a vector because the activation function is vector function and two indices need to be accounted for.

like the sigmoid function (see figure 6.4). A good choice to avoid this scenario, and the most widely used one, is the ReLU function (Rectifier Linear Unit) defined as $g(z)_{relu} = \max(0, z)$. This means that the function returns the input value if the input value is positive and it returns a zero if the input is negative. It is computationally cheap to use and converges fast: since the function doesn't plateau, not only are vanishing gradients rare, but the learning is also fast. Finally, it is sparsely activated: since the output is often zero it is not likely for a unit to be activated thus reducing overfitting problems. The output layer has to be adapted to the problem at hand. For a binary classification problem the sigmoid function is the default one and its shape can be seen in figure 6.4. It is defined as:

$$g(z)_{sig} = \frac{1}{1 + e^{-z}} \quad (6.3)$$

What the sigmoid function does is split the examples between those that activate the output layer and those that don't thus creating a separation/classification.

For multiclass classification the softmax activation function is used in the output layer. Softmax can be thought of as a softer version of the argmax function. The argmax function returns the index of the largest value in a list: if the inputs to the last layer activation function is $\vec{z}_{in} = [1, 3, 2]$ the argmax function will return $\vec{z}_{out} = [0, 1, 0]$. We can create a function that instead of returning such a hard output gives a probabilistic interpretation of what the max index value is. The probability for each component z_j of the input vector \vec{z}_{in} to be the max value is then calculated as:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{j=1}^K e^{z_j}} \quad (6.4)$$

Applied to the input vector this returns $\vec{z}_{out} = [0.09, 0.67, 0.24]$. To use the softmax activation function in a classification problem, the labels of the classes need to be one-hot encoded, meaning that we need a probabilistic interpretation of our classes as well to be able to confront them with the output of the network. Three classes 0, 1 and 2, for instance, will be one-hot encoded as $[1, 0, 0]$, $[0, 1, 0]$ and $[0, 0, 1]$.

6.2.4 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are a type of DNN and, as such, are made of neurons with weights and biases. The main difference amongst the simple configuration described in the previous paragraph and CNNs, is the shape of the input and the way it is treated. Inputs to CNNs consist of three dimensional matrices instead of vectors, in general they are used to process multi-channelled images (RGB most of the times). Why the need for a different treatment of the data? The number of parameters in a network and, as a consequence, how computationally intensive its optimisation is, can grow very fast: the addition of a 1000-unit layer means adding 1 million parameters to learn [102]. If we consider all pixels within RGB images of size 100x100 as features, we have 30000 input features per image. If a DNN were used to solve a classification problem with this many inputs it would likely never converge. A way to efficiently extract relevant features in the matrix needs to be found.

CNNs were developed with two ideas in mind, the first is that nearby pixels (or points in a matrix) are more related to each other than far away ones and the second one is that different objects within an image are separated by edges. We need a method that allows the model to learn both.

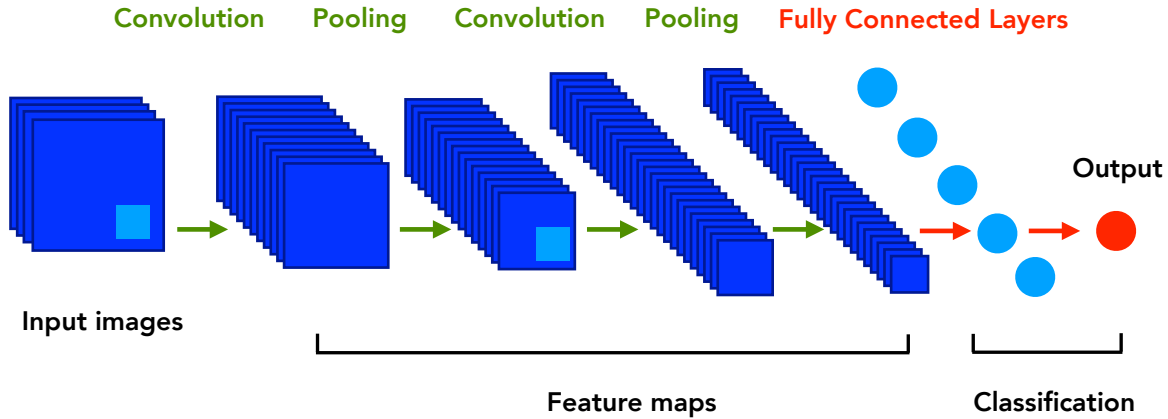


Figure 6.5: Typical architecture of a simple convolutional neural network.

To account for the first point, a filter F , a small squared matrix with randomly initialised values, slides over the full image, divided in fixed size patches P , and convolves it. The step size of the sliding window is referred to as the stride and is defined before hand by the user. A convolution is the scalar product between the filter and the image patches [107]. The more similar the filter is to the patch the higher the result of the dot product. A bias parameter is also added after the convolution and before the activation function [102]. At the end of a convolution over the entire image we are left with a smaller matrix that should represent a condensed version of the image. In each layer of the CNN the convolution of the image is done using several, n , different filters. This means that after each CNN layer we have n different condensed representations of the image, also called feature maps. The elements of the filters and the bias associated are the parameters that need to be learned by the algorithm. As with DNNs, an activation function is applied to introduce non-linearity after each layer. When more layers are added each new layer takes as inputs smaller versions of the original image. Another building block of convolutional neural networks are pooling layers. These layers usually follow the convolution layers and reduce the number of parameters to learn. As for convolution, pooling layers have a stride and a filter size, however, in this case the filter consists of a fixed operator and not in something to learn. Some widely used pooling operators are the max function and the average/mean function.

Once the feature maps have been learned, fully connected layers of the type described in the DNN section 6.2.2, are applied. The classification process is therefore done using the convoluted features as inputs. A typical architecture for a CNN can be seen in figure 6.5.

Direct application of CNNs have been tried in particle physics by building jet “images” [108, 109], the idea is that each constituent is located in grid using its angular coordinates and the channels of the image represent the various features of interest (such as the transverse momentum or the energy).

6.2.5 Hyperparameters and Regularisation

Good choices in hyperparameters determine the success of a learning algorithm to solve the problem at hand. Along our description of how neural networks work we have already introduced several hyperparameters: number of epochs, stride, filter size, choice of pooling function, choice of the parameter initialiser and the choice of the activation function for each layer. Several others exist. In particular the batch size (BS) controls the number of training samples to work through before the model parameters are updated. Increasing this number (default is $BS = 1$) allows the training to be split over a larger number of examples and reduce the number of times the model is evaluated.

We have touched on learning rate, however it should be noted that determining the optimal learning rate is crucial: too small a learning rate means that the network might not converge, instead, if the value is too large, the loss can fluctuate around a local minimum.

The choice of the optimisation strategy can also be seen as an hyperparameter. We have talked about gradient descent but many optimisation techniques have been developed starting from this principle. A particularly useful optimiser is Adam [110]. Adam is a method for stochastic optimisation, it uses different learning rates for different parameters that need to be learned and automatically adapts the values of the learning rates during training. It is especially useful when large datasets are used or when the model is comprised of a large number of parameters to find.

Finally, regularisation terms are hyperparameters needed to reduce overfitting: two of the most widely used methods are **L1 regularisation** and **L2 regularisation**. Both L1 and L2 add a penalty term, that depends on the parameter value of the input features, to the loss function. In practice, L1 regularisation creates a sparse model by “turning off” most of the features and therefore performing a sort of feature selection. L2 is differentiable and generally performs better if there is no need to reduce the number of features but the only goal is to improve performance on the testing data.

Other regularisation techniques can also be used:

- Batch normalisation [111] applies standardisation procedures to the outputs of each layer. It improves the performance of the model and increases the convergence speed. Even though it wasn’t developed as a regularisation procedure it also has a regularising effect.
- Dropout consists of a very simple concept: the idea is to randomly and temporarily exclude some units of the network each time it evaluates a training example. This increases the sparsity of the network and makes it more robust against overfitting.
- Early stopping is the practice in which the model is saved with a given frequency and the performance of each of these preliminary networks is assessed over the validation dataset. In the case of overfitting the loss function computed over the training sample will get smaller and smaller whereas when evaluated on the validation dataset it will stop decreasing and, in some

cases, will get worse. Checking how the model performs on unseen data allows us to stop the training before severe overfitting occurs.

6.3 Edge Convolution

Edge convolution (EdgeConv) is a new neural network configuration that builds on CNNs. It is designed to be applied to point cloud inputs [112]. Point clouds are a representation of three dimensional objects that are widely used in computer graphics. Each object is represented by an unordered and irregularly distributed set of points defined by their positional coordinates and features. In particle physics jets can be seen as an unordered, permutation invariant set of particles (constituents) or, as we will refer to them in the rest of the thesis, **particle clouds** [113]. On top of the irregular distribution in 3D space, particle clouds and point clouds share other similarities. In particular, since they make up higher level objects (jets, 3D shapes) that have a rich internal structure, the points/particles are far from being unrelated amongst themselves [113]. This suggests that an architecture that is able to perform classification tasks on point clouds is likely to perform well on particle clouds.

What makes convolutional neural networks so successful for computer vision tasks is a combination of how the image inputs are processed and the ability to stack various convolutional layers (see section 6.2.4). Indeed CNNs exploit the translational symmetry of images to learn using a sliding window technique that, on one hand reduces the dimensionality of the inputs and, on the other, also guarantees efficient learning of important features since all image locations are used to determine the parameters of the model. The stacking of convolutional operations allows for hierarchical learning: shallow layers will learn directly from the image patches and will create low-level feature maps, deeper layers can exploit this information and create more complex representations. To extend the CNN approach to point clouds both aspects need to be reproduced in some way.

The problem with point clouds is that they are irregularly distributed meaning that the idea of convolution on fixed size patches cannot be replicated as is. On top of that, even if the inputs were to be split in fixed patches the invariance under permutation, which is a key property of point clouds, wouldn't be preserved. EdgeConvs solve this problem by creating a local neighbourhood graph onto which convolution operations can be applied [112]. The vertices of the graph are the points in the point cloud, the edges of the graphs are the connections between each point and its k nearest neighbours, also called edge features. Edge features are defined as:

$$e_{ij} = h_{\Theta}(x_i, x_j) \quad (6.5)$$

where h_{Θ} is a non linear function with a set of learnable parameters Θ . The convolution operation is done over the edge features by applying a channel-wise symmetric aggregation operation \square (such as max or average). The output of the EdgeConv operation at the i_{th} vertex is therefore given by:

$$x'_i = \square_{j=i}^k h_{\Theta}(x_i, x_j) \quad (6.6)$$

The application of the EdgeConv operation can be seen in figure 6.6.

The choice of the \square aggregation function and the h_{Θ} edge function is decisive in the performance of the network. If the edge function is chosen to be a simple matrix operator we go back to standard

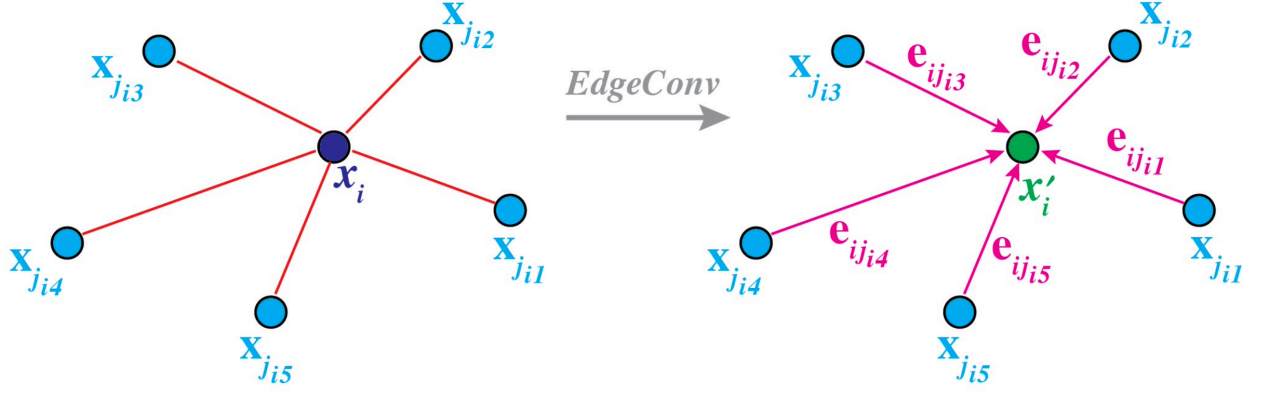


Figure 6.6: The EdgeConv operation for x_i . The output x'_i is the result of the aggregation of all the edge features computed. Image from the EdgeConv paper [112].

convolution over a patch with centre x_i and made up of k “pixels”. To capture both global information, in the form of the coordinates of the patch centre, and local neighbourhood information, captured by the differences between the centre and all the neighbourhood points $x_i - x_j$ the choice for the edge feature function falls onto:

$$h_{\Theta}(x_i, x_j) = \bar{h}_{\Theta}(x_i, x_i - x_j) \quad (6.7)$$

In particular, the \bar{h}_{Θ} function can be implemented as a multilayer perceptron whose parameters are shared among all edges, as can be seen in figure 6.7 for two points defined only by their 3D coordinates.

The aggregation function is commonly chosen to be either the max function, the average function or the sum function. For most vision applications the max function is the one that grants the best results, previous tests on particle physics data, however, show improved performance using the average function. One key aspect of edge convolution is the possibility of stacking several EdgeConv operations thus creating the same hierarchical learning of CNNs. In particular, graphs built using EdgeConv are dynamic, meaning that the set of nearest neighbours changes from layer to layer. Proximity in feature space differs from proximity in the original point cloud [112], so that, in practice, the edge features learned during the EdgeConv process can be seen as new coordinates of the original point in a latent space [113]. It should be noted that finding nearest neighbours implies the use of a distance metric (hyperparameter) to determine which points are close to each other and which ones are far apart. For most problems the typical euclidian distance can be used but other distance definitions can also be taken into consideration (especially for particle physics applications).

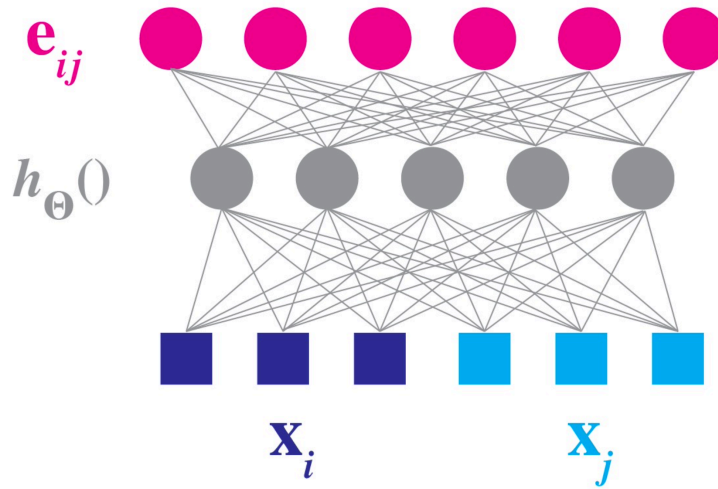


Figure 6.7: Computation of a single edge feature from a pair of points. Image from the EdgeConv paper [112].

6.4 Software

Keras is a python software library dedicated to neural networks and the one used to develop the classifiers presented in the next chapter. It can be run on top of several machine learning platforms but, in the context of this thesis, the Tensorflow backend is used [114]. Unlike most ATLAS software both Keras and Tensorflow are developed in python, meaning that some translation between the machine learning and physics frameworks is necessary both in terms of data preprocessing and deployment of trained algorithms.

[10pt]article [usenames]color amssymb amsmath [utf8]inputenc

Improving boson tagging abilities

Contents

| | | |
|------------|--|------------|
| 7.1 | Polarisation from an experimental point of view | 122 |
| 7.2 | Samples and datasets | 126 |
| 7.3 | DNNs | 127 |
| 7.4 | EdgeConv | 131 |
| 7.5 | Combined | 138 |
| 7.6 | Outlook | 143 |

In the previous two chapters we have described both the details and importance of the vector boson tagger within the VVJJ analysis, and the concepts of machine learning based classification. The aim of this chapter is to combine the two and create a machine-learning-based tagger for an improved future iteration of the VVJJ analysis. The use of machine learning for classification tasks also opens the possibility to build new taggers that do not only discriminate between QCD and vector boson jets, but also separate jets originating from W bosons from those originating from Z bosons, or which are even able to identify the polarisation of said bosons.

The reasons behind the choice of using machine learning based algorithms are many. First, ML techniques allow us to exploit more of the available information: while algorithmic strategies that scan all possible cut combinations (as for the tagger described in chapter 5) are limited to scanning a 2 or 3 dimensional variable space, DNNs instead can take in as many needed inputs and automatically weigh them to create a model that best maps them to the correct class. Second, neural networks such as Edge Convolutions can be used directly on jet constituents in order to automatically extract features from low level data. This feature-engineering aspect of machine learning could allow us to extract relevant information not encoded in the form of any currently existing jet substructure variable.

It is this last point, in particular, that is very interesting when building a tagger that goes beyond the task of QCD rejection. Indeed, even though many jet properties can be inferred from the currently available jet substructure variables (for instance D_2 is a measure of how compatible a jet is with a two-prong structure), others, such as polarisation and boson nature⁽¹⁾, remain unexplored.

The ability to differentiate between W and Z jets and to identify their polarisation could prove critical

⁽¹⁾With the exception of mass which is a JSS variable that provides some discrimination between W and Z jets.

in a future iteration of the VVJJ analysis for various reasons. First, if the search were indeed successful and a new particle were to be found, knowing the nature of the bosons involved would allow us to know with which class of theory the observation is compatible. Second, di-boson interactions with the addition of polarisation and/or nature information could highlight deviations from precise Standard Model expectations.

The existence of the Higgs boson, described in chapter 1, can also be derived from the principle of unitarity. Indeed, the scattering of same sign longitudinally polarised W bosons ($W_L^\pm W_L^\pm \rightarrow W_L^\pm W_L^\pm$) violates unitarity since the amplitude of the phenomenon diverges with the second power of energy. The problem can be solved by introducing a new particle such as the Higgs boson, which produces amplitude counter-terms that cancel the divergences [115]. In particular, it must be that the coupling g_{HWW} of the Higgs boson exchanged between two W bosons exactly satisfies the following condition:

$$g_{HWW} = gM_W. \quad (7.1)$$

Therefore, by introducing a Higgs boson with the appropriate couplings to other particles, the unitarity problem is solved. However, for compatibility with the Standard Model, the Higgs boson must have a mass that allows it to be produced before the energy scale of unitarity violation: if the Higgs boson is too heavy, the unitarity problem remains. The mass constraint derived from di-boson scattering is one of the reasons behind the choice of energy scale of the LHC; indeed, since longitudinal W boson scattering had been observed, some mitigating process had to happen before a few TeV. The discovery of the Higgs boson in 2012 proved its existence but many questions still remain: is the Higgs boson an elementary particle, what is the origin of its mass (hierarchy problem), etc... Deviations from the precise Standard Model expectations for di-boson interaction amplitudes offer a prime probe for new physics but can only be tested if the polarisation and the nature of the vector bosons can be identified, as can be seen in figure 7.1.

In short, we want to be able to separate QCD jets from vector bosons jets (QCD rejection tagger - QCDRej), W jets from Z jets (vector boson nature tagger - VBN) and longitudinally polarised jets from transversely polarised jets (polarisation tagger - Pol). These three tasks describe three binary taggers (a classification algorithm that needs to separate between two classes), each of which will be developed using three different machine learning algorithms: Deep Neural Networks (section 7.3), EdgeConv (see section 7.4) and a combination of the two (section 7.5). A summary of the three tagger types and the classes used as signal and background can be found in table 7.1.

The goal of this chapter is to demonstrate the feasibility of each task and to compare the performances achieved using different techniques. Since the ability to separate between vector bosons and QCD is similar for W s and Z s (see chapter 5) and we want to separate the impact of polarisation, the QCD rejection tagger is trained to separate longitudinally polarised W bosons from QCD. This also allows for comparisons with past QCDRej taggers, which were optimised under the same strategy.

7.1 Polarisation from an experimental point of view

In practice, as discussed in section 1.3.3, polarisation results in a preferred spin direction and can be quantified through the use of helicity h (the projection of the particle's spin S onto the direction of

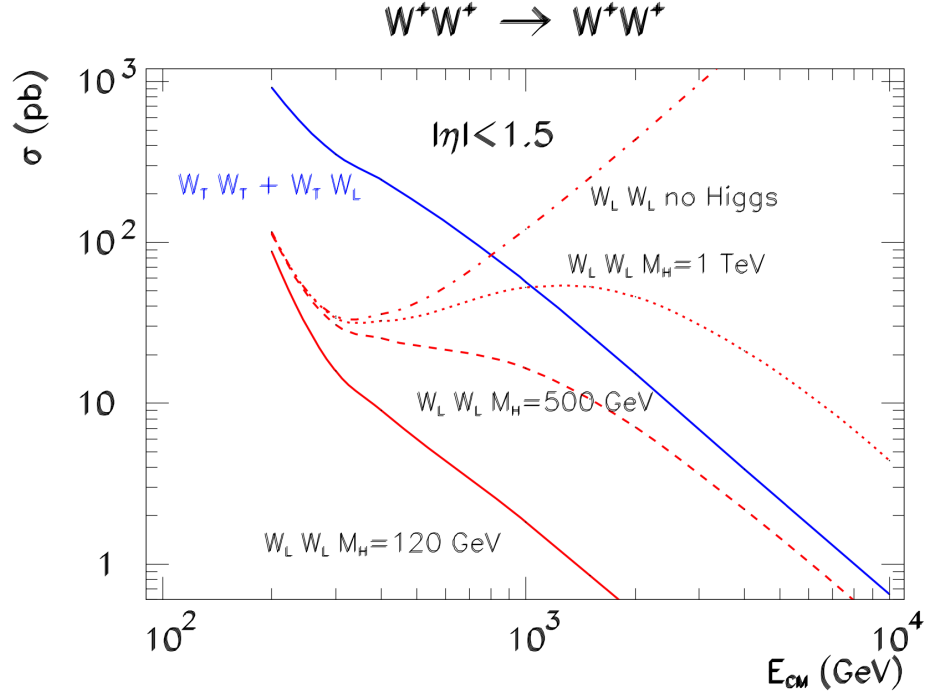


Figure 7.1: The total W^+W^+ scattering cross sections as a function of the centre of mass energy for different final (and initial) state polarisations and for different Higgs masses, including the limiting Higgsless case. Assumed are two on-shell, unpolarised, colliding W^+ beams. A cut on the scattering angle that corresponds to pseudorapidity of ± 1.5 with respect to the incoming W direction was applied. The individual $W_T W_T + W_T W_L$ curves for each Higgs mass value coincide within the width of the blue line. The results were the product of MadGraph calculations [115].

| Tagger type | Signal class | Background class |
|-------------|--|--|
| QCDRej | Longitudinally polarised jets originating from W bosons (WL jets) | QCD jets |
| VBN tagger | Jets originating from Z bosons with mixed polarisation (Z jets) | Jets originating from W bosons with mixed polarisation (W jets) |
| Pol tagger | Longitudinally polarised jets originating from either W or Z bosons (VL jets) | Transversely polarised jets originating from either W or Z bosons (VT jets) |

Table 7.1: Summary of the types of taggers optimised. The signal and background classes are specified for each.

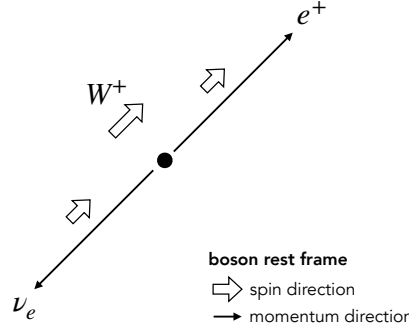


Figure 7.2: Directions of momenta and spins of the decay products of a W^+ in its rest frame.

its momentum \vec{p}):

$$h = S \cdot \frac{\vec{p}}{|\vec{p}|} \quad (7.2)$$

Since we are working with vector bosons, i.e. spin 1 particles, only three possible helicity values are allowed : -1, 1 and 0. A state with $h = \pm 1$ is transversally polarised whereas $h = 0$ corresponds to a longitudinal polarisation.

Spin conservation dictates relations between the helicity of a particle and the angular distribution of its decay products [116]. For simplicity, let's consider a W boson (which can only couple to left-handed particles and right-handed antiparticles): in its rest frame, the charged lepton will be emitted in the direction of the spin for W^+ and in the opposite direction for W^- (see figure 7.2).

This emission direction can be quantified by the decay angle $\theta_{W/Z}^*$ between the direction of the momentum of the boson and the charged lepton's momentum in the W/Z rest frame (see figure 7.3). The distributions of the $\theta_{W/Z}^*$ angles can be seen in figure 7.4. We can see that, generally, the decay products of longitudinally polarised vector bosons tend to decay perpendicularly with respect to the direction of the parent vector bosons whereas for transversely polarised vector bosons the opposite is true.

The angle θ_X between the beam axis and the vector bosons is also related to the polarisation of the vector bosons, as can be seen in figure 7.5. The distribution of this angle, however, is process dependent, meaning that the distribution will change not only based the polarisation of the vector bosons, but also on the nature of the primary collision (quark-quark, gluon-gluon or quark-gluon) and the decay mode of the hypothetical new particle.

In the context of the VVJJ analysis the θ_X angle, being process dependent, can be used as a final state selection but is not generic and requires a specific interpretation. The θ_V^* angle provides jet level information about polarisation and has been extensively studied in the context of vector bosons

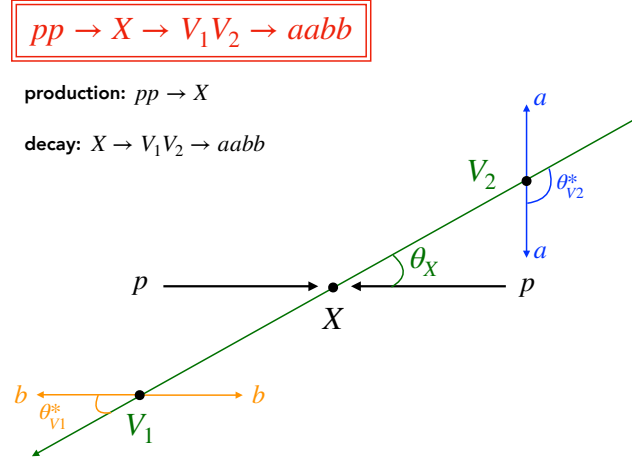


Figure 7.3: Definition of the decay angle $\theta_{V_{1,2}}^*$ and the angle between the vector bosons and the beam axis, θ_X .

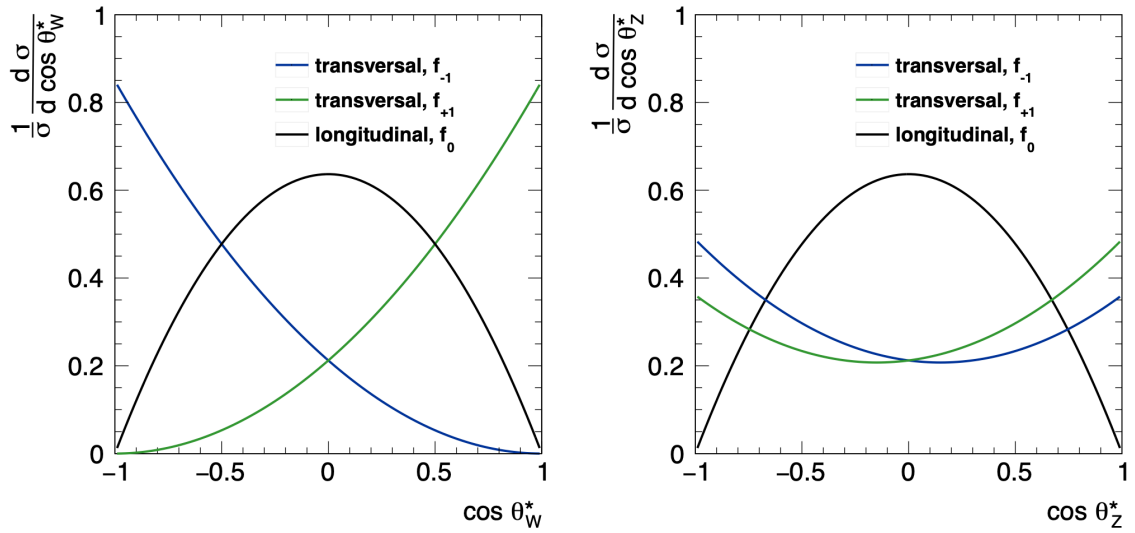


Figure 7.4: Normalised differential distributions of the cross section over the cosine of the decay angle of a W^+ boson, (a), and of a Z boson, (b). In blue and green the contributions f_{-1} and f_{+1} of the transversely polarised bosons to the total cross section and in black the contribution f_0 of the longitudinally polarised bosons [116]. The details of the calculations of the cross section and of the fractions can be found in [117].

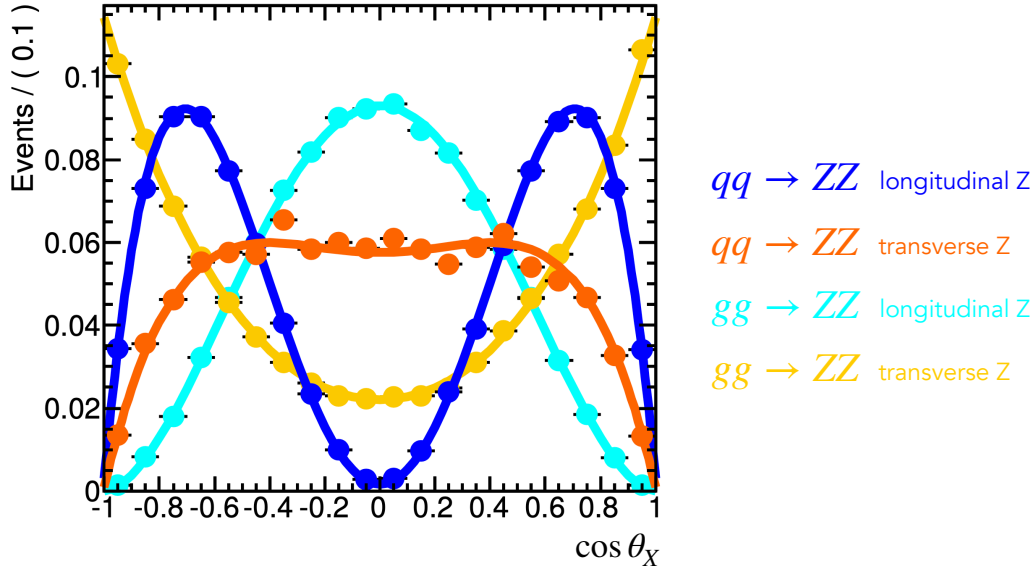


Figure 7.5: The angle between the vector bosons and the beam axis [118].

decaying leptonically, in which it can be precisely measured [116, 119], and could also potentially be used for fully hadronic decays at low momentum (where the two quarks can be reconstructed as two distinct small-R jets). At high p_T , however, the decay products of the boosted vector bosons are highly collimated, making the measurement of the angle between them, and their Lorentz transformation back to the boson’s rest frame⁽²⁾, quite challenging. Some JSS variables might offer some sensitivity to this angle: Split12, τ_{21} , Z_{12}^{cut} and PtImbalance (see section 4.6) show some separation between different polarisations. As already mentioned in section 4.6, the use of multiple variables does not always result in more discrimination power due to the correlations between them. As such, the possibility of extracting different and relevant information directly from the jet constituents, provided by the EdgeConv architecture, and combining it to the already known JSS variables might prove useful.

7.2 Samples and datasets

All the taggers presented in the rest of the chapter are derived using MC simulated proton-proton collision events at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. All simulated events are passed through GEANT4 [120] which provides a complete ATLAS detector simulation. In particular, two sets of samples providing hadronically decaying W and Z bosons (one for longitudinal polarisation and one for transverse polarisation) and one set of multijet (QCD) samples were used.

Multijet production is used to define the QCD background originating from light quarks and gluons. These processes are generated using Pythia v8.230 [79], the NNPDF2.3LO PDF set [80] and the A14 set of tuned parameters [81].

⁽²⁾The helicity is not Lorentz invariant meaning that in the laboratory frame the directions of the decay products of boosted vector bosons will change.

For the two nominal W/Z boson signal samples, high-mass SSM $W' \rightarrow WZ \rightarrow qqqq$ events are simulated using Pythia v8.235 [79] with the NNPDF2.3LO PDF set [80] and the A14 set of tuned parameters [81] and filters are applied to select the polarisation of the vector bosons [82].

The first step in building a Machine Learning based tagger, as detailed in chapter 6, is to prepare the datasets. For all three tasks, the examples within the datasets consist of large $R = 1.0$ anti-kt UFO jets with softdrop grooming and CS+SK pile-up mitigation, as ATLAS has identified to be optimal [65]. In the end of the chapter, performance comparisons to taggers developed using trimmed large $R = 1.0$ TCC jets are also shown.

The jets in the samples described above are randomly split into testing, training and validation datasets. The testing dataset is common to all three typologies of taggers and comprises jets with transverse momentum ranging from 0.4 to 2.5 TeV. Specifically, it is made up of 5×10^5 QCD jets weighted to produce the expected smoothly falling jet p_T distribution of the multijet background, 5×10^5 jets originating from W bosons (50% longitudinally polarised and 50% transversely polarised) and 5×10^5 jets originating from Z bosons (50% longitudinally polarised and 50% transversely polarised).

Since the number of jets per class needs to be balanced (section 6.1), the training and validation datasets are different for the QCD rejection tagger and the boson nature/polarisation taggers. For QCDRej, the training sample comprises of 1.2 million jets, half of which are longitudinally polarised jets originating from W bosons and half of which are QCD background jets. The polarisation tagger and the VBN tagger are trained on the same data: 1.2 million jets, half of which are W jets (50% longitudinally polarised and 50% transversely polarised) while the remaining half are Z jets (50% longitudinally polarised and 50% transversely polarised). The training and validation datasets are constructed in the same way (same proportion of examples per class) but differ in size: the validation dataset (1.2×10^5 jets) is ten times smaller than the training dataset.

Finally, it should be noted that although the datasets for each tagger type contain the same events for all three architectures, the input features vary between them. The number and nature of said inputs will be detailed for each separate architecture in the rest of the chapter.

7.3 DNNs

The DNN taggers are obtained by training fully connected feed-forward neural networks with multiple dense⁽³⁾ hidden layers for 200 epochs with early-stopping. The performance of this type of network is entirely determined by the choice of hyper-parameters and input features (in this case, a combination of JSS variables). As such, several combinations of input features and hyper-parameters are tested for each of the tagger types. Some hyper-parameters can, however, be fixed in advance based on literature results. In particular: the optimiser is chosen to be the Adam algorithm due to its generalised success in the context of classification tasks [110], the weights initialiser is glorot-uniform [121], the activation function of the hidden layers is the ReLu function [122] and, due to its applicability to binary classification, the activation function of the output layer is a sigmoid.

⁽³⁾A dense layer is a layer that is fully connected to all the nodes of the previous layer.

| Tagger type | Input groups | Inputs |
|---------------|--------------|---|
| QCDRej tagger | nJSS = 18 | $p_T, m, ECF3, \tau_{21}, Split12, C2, D2, Angularity, Aplanarity, R2FW, P, KtDR, \tau_1, \tau_2, zcut12, nConst, ug_{nTrk}, PtImbalance$ |
| VBN tagger | nJSS = 1 | m |
| | nJSS = 4 | $KtDR, Split12, R2FW, m$ |
| | nJSS = 6 | $KtDR, Split12, R2FW, m, P, PtImbalance$ |
| | nJSS = 18 | $p_T, m, ECF3, \tau_{21}, Split12, C2, D2, Angularity, Aplanarity, R2FW, P, KtDR, \tau_1, \tau_2, zcut12, nConst, ug_{nTrk}, PtImbalance$ |
| Pol tagger | nJSS = 1 | $zcut12$ |
| | nJSS = 4 | $zcut12, Split12, \tau_{21}, D2, PtImbalance, Angularity$ |
| | nJSS = 6 | $zcut12, Split12, \tau_{21}, D2$ |
| | nJSS = 18 | $p_T, m, ECF3, \tau_{21}, Split12, C2, D2, Angularity, Aplanarity, R2FW, P, KtDR, \tau_1, \tau_2, zcut12, nConst, ug_{nTrk}, PtImbalance$ |

Table 7.2: Jet substructure variable groups tested during the optimisation of the QCDRej tagger, the Pol tagger and the VBN tagger.

The list of input variables for the QCD rejection tagger is taken from previous extensive tests realised in the context of the search for pair production of heavy vector-like quarks in hadronic final states [123]. For VBN and polarisation tagging, three input variable groups are tested (the groups for each tagger type can be seen in table 7.2). The groups are constructed based on the observation of the histograms of the JSS variables for each class: one group consists of the four variables that offer the largest separation between the classes and for which the separation is independent from p_T , a second group adds two other variables which offer separation in specific p_T regions, finally a third group contains the same variables used for the QCD rejection tagger for consistency. For the polarisation tagger, the performance of the DNN tagger is also compared to that of two taggers that take as input solely the $zcut12$ variable: a simple cut based tagger on $zcut12$ which is the variable that offers the most discrimination (denoted “ $zcut12$ cut” in the figures) and a DNN tagger with single layer and a single node (denoted “ $zcut12$ DNN” in the figures). For the vector boson nature tagger, the performance of the DNN tagger is also compared to that of a simple cut based tagger on the mass of the vector bosons (denoted “ m cut” in the figures).

The output of a neural network architecture is a single discriminant that allows for the classification of a jet as either signal-like (WL jet, Z jet and VL jet respectively for the three tagger types) or background-like (QCD jet, W jet, VT jet respectively for the three tagger types). The distributions of the DNN discriminants for the final configuration of the three tagger types are shown in figure 7.6 in a

| Number of hidden layers | Number of nodes per hidden layer |
|-------------------------|----------------------------------|
| 1 | 18 |
| 3 | 27, 18, 9 |
| 5 | 25, 22, 20, 14, 7 |
| 7 | 36, 32, 29, 25, 22, 18, 14 |

Table 7.3: Table indicating the various DNN architectures tested during the optimisation of the QCD Rej, Pol and VBN taggers. Both the number of layers and the number of nodes per layer are indicated.

single transverse momentum bin⁽⁴⁾. The working points of the taggers are defined as single-sided cuts that are a function of the reconstructed jet p_T so that they yield a constant signal efficiency (50%) in each p_T bin⁽⁵⁾. As such, a jet in the testing dataset is considered as tagged if its discriminant score is higher than the cut corresponding to its p_T bin.

During the optimisation procedure, the performance of the various DNNs is characterised by measuring the relative signal efficiency and relative background rejection as a function of transverse momentum. These quantities are evaluated using the jets from the testing datasets. The relative signal efficiency for each p_T bin i is defined as:

$$\varepsilon_{sig,i}^{rel} = \frac{N_{sig,i}^{tagged}}{N_{sig,i}^{total}} \quad (7.3)$$

Similarly, the relative background rejection for each p_T bin i is defined as $1/\varepsilon_{bkg,i}^{rel}$.

The results for the best group of variables to be used for the VBN tagger and the Pol tagger are shown in figure 7.7. Given that the nJSS = 18 group yields the highest (VBN) or same (Pol) background rejection in all p_T bins compared with other input choices, the same set of input variables is used for the three DNN tagger types during the hyper-parameter optimisation procedure and in the final DNN configurations.

The results of the optimisation for the hyper-parameters can be seen in figure 7.8 for the choice of batch size, in figure 7.9 for the choice of the learning rate, in figure 7.10 for the choice of the regularisation constant $l1$ and in figure 7.11 for the number of layers (table 7.3 shows the number of nodes per layer for the configurations tested).

Since the hyper-parameter values that give the best background rejection are comparable or the same for all three tagger types, the same feed forward neural network architecture is used in all three cases. In particular, we use the same 18 JSS variables as inputs, a batch size of 2000, a learning rate $lr = 1e^{-5}$, no $l1$ or $l2$ regularisation and a network built out of five dense hidden layers.

⁽⁴⁾The discriminant distributions for the EdgeConv and Combined tagger presented later are not shown since the performance of the architecture is better represented in the form of the background rejection.

⁽⁵⁾This procedure is the same done for the cut-based tagger described in chapter 5 but instead of finding the cuts that yield maximum significance, we look for the cuts that give a fixed efficiency value.

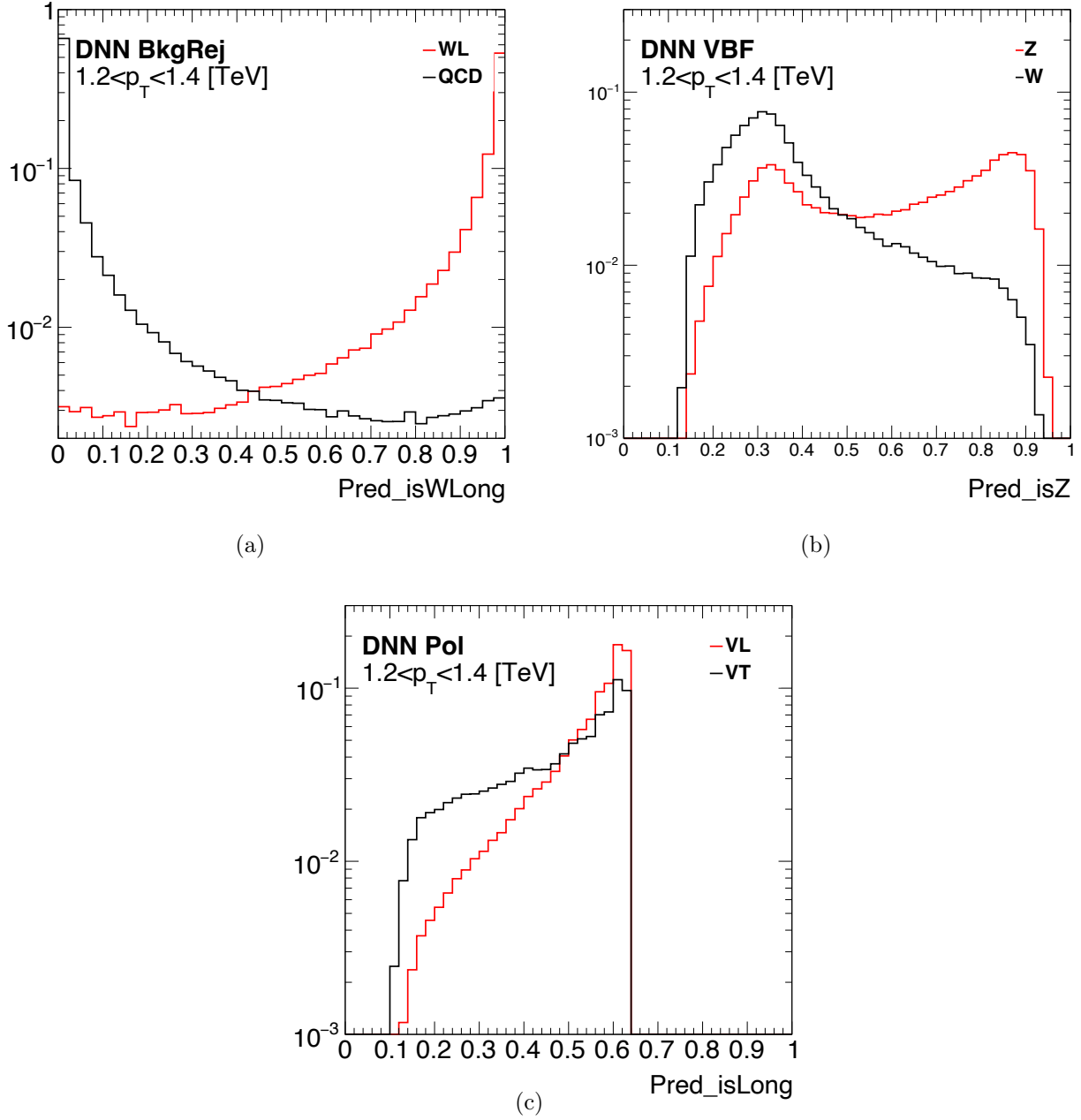


Figure 7.6: Discriminant distributions for the final configuration of (a) DNN QCDRej tagger, (b) DNN VBN tagger and (c) DNN Pol tagger. The legend denotes the truth label.

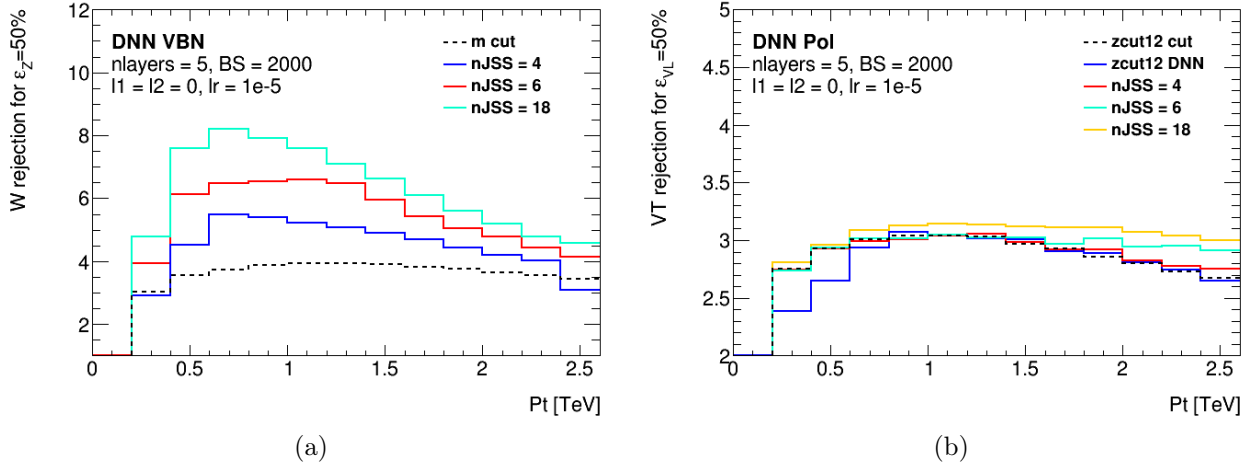


Figure 7.7: Background rejection as a function of transverse momentum for different input variables to (a) DNN VBN tagger and (b) DNN Pol tagger. The $nJSS = 18$ group gives best results in both cases. The optimal number of input variables for the QCDRej tagger is obtained from previous studies [123]

7.4 EdgeConv

The Edge Convolution network described in section 6.3 can be seen as an algorithmic block, or layer, which can be encased within a more general neural network architecture. Four different architectures have been tested to find the one that gives the best performance. The differences between the architectures can be found both in the EdgeConv blocks and in the overall architecture. A schema of all five architectures can be seen in figure 7.12.

In particular there is the **EC Baseline** architecture, which is comprised of two EdgeConv layers, within which a single convolution layer (CNN) with a small number of filters is present, and three Dense fully connected layers that allow the model to map the features extracted by the EdgeConv layers to the desired output. The **EC MultConv(MC)** network is characterised by three EdgeConv layers that, instead of one, are made up of three CNN layers with a large number of filters; the rest of the model architecture is the same as **EC Baseline**. The **EC MC+BN+Relu** builds on the **EC MultConv(MC)** model but adds an activation ReLu function and a Batch Normalisation function after each CNN layer within the EdgeConv block. The **EC MC+BN+Relu+Dropout** is the same as the one just described with the exception that Dropout layers are added after each Dense layer (the two architectures are shown together for space constraints). Finally, the **EC ParticleNet** copies the architecture presented in the article relating the first application of Edge Convolution to particle physics problems [113].

The inputs to the model are fixed-size tensors⁽⁶⁾ of shape $(n_{Jets}, n_{Constituents}, n_{Features})$. In reality, however, the number of constituents varies from jet to jet, hence a padding strategy needs to be applied. Two, in particular, were tested:

⁽⁶⁾This is due to the presence of at least one CNN layer within the algorithmic block (see section 6.1.2).

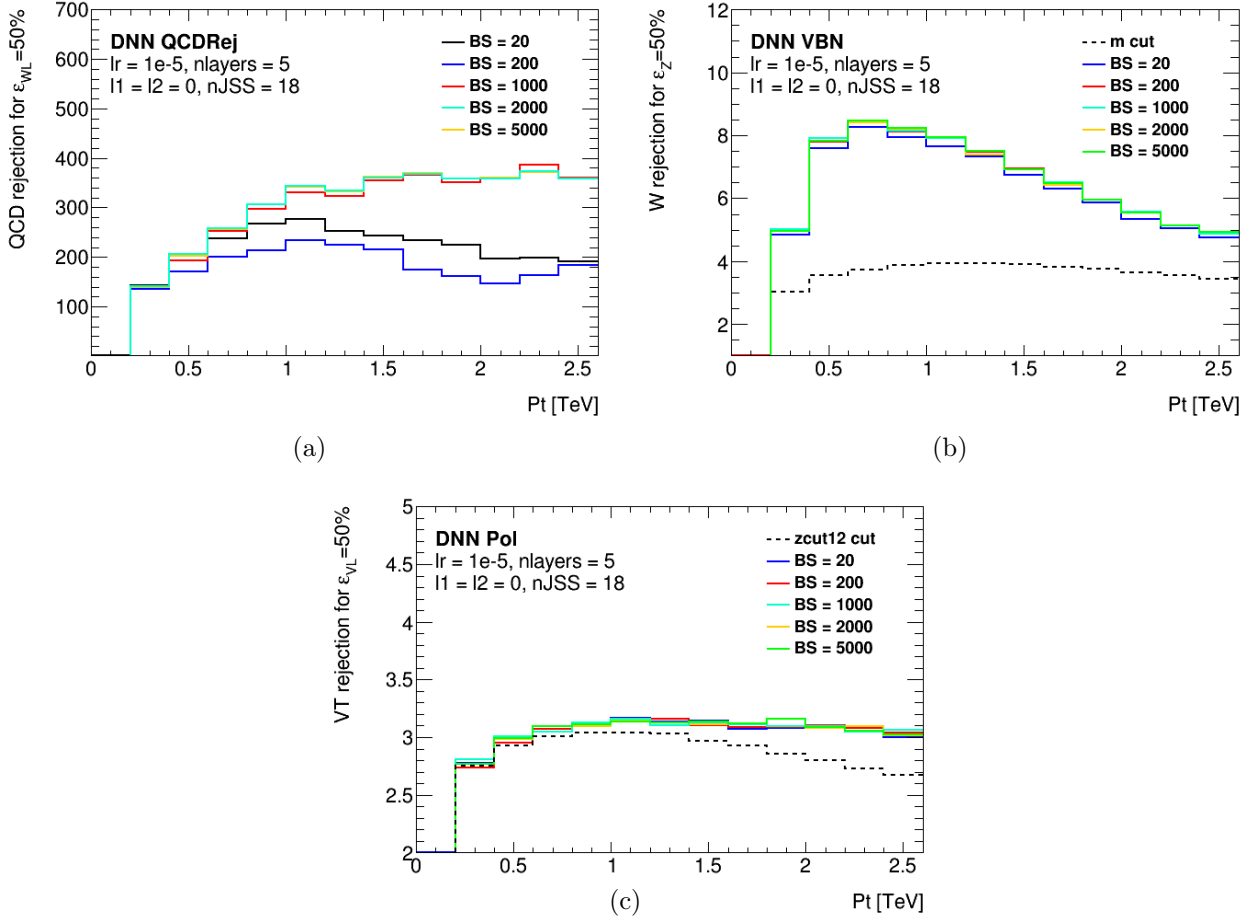


Figure 7.8: Background rejection as a function of transverse momentum for different batch size (BS) values for (a) DNN QCDRej tagger, (b) DNN VBN tagger and (c) DNN Pol tagger. The results for batch size between 1000 and 5000 are consistently the best for all three tagger types without much separation between them, as such a value of $\text{BS} = 2000$ is chosen for the final architecture.

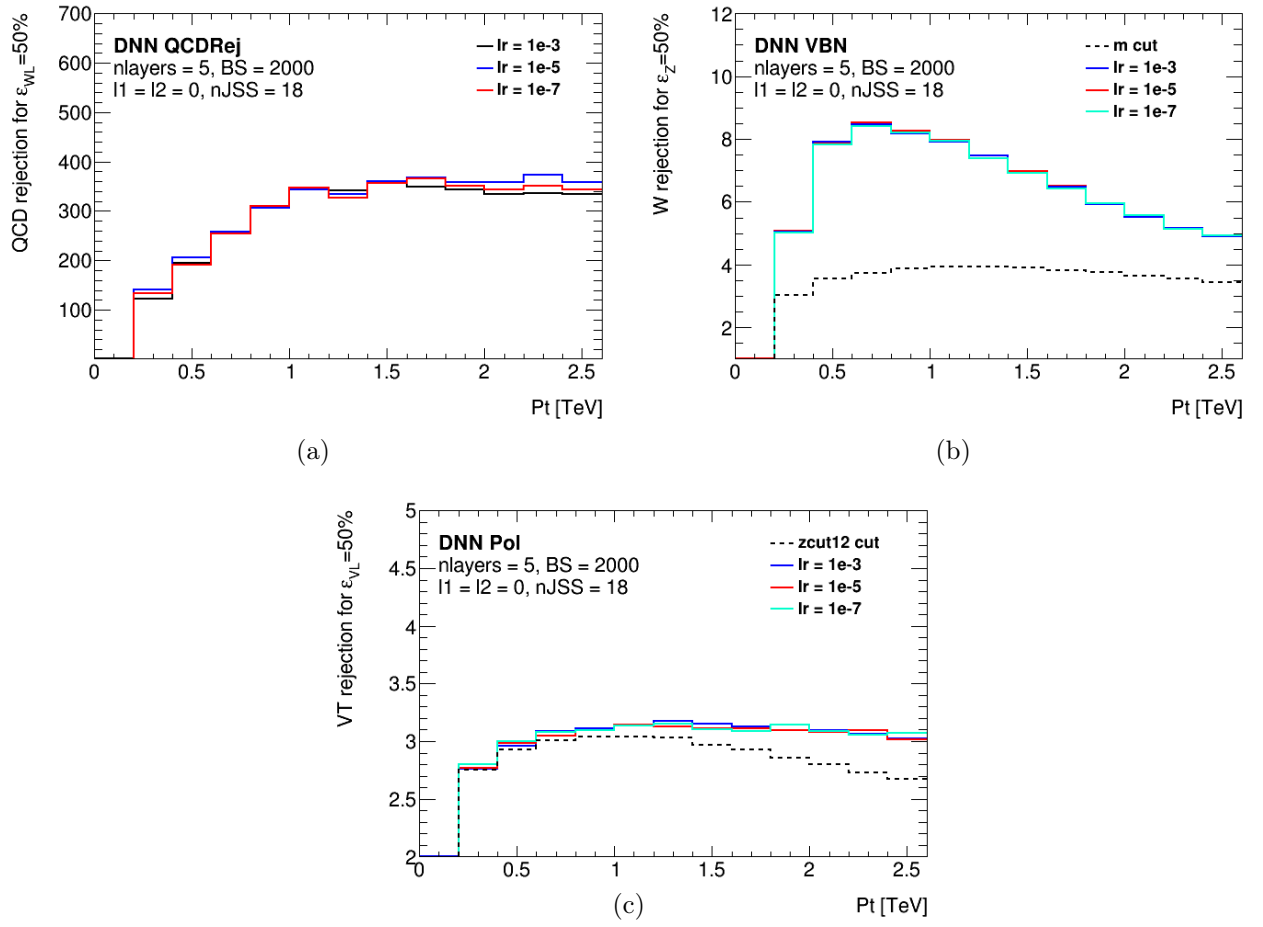


Figure 7.9: Background rejection as a function of transverse momentum for different learning rate (lr) values for (a) DNN QCDRej tagger, (b) DNN VBN tagger and (c) DNN Pol tagger. A learning rate $lr = 1e-5$ yields good background rejection for all three tagger types.

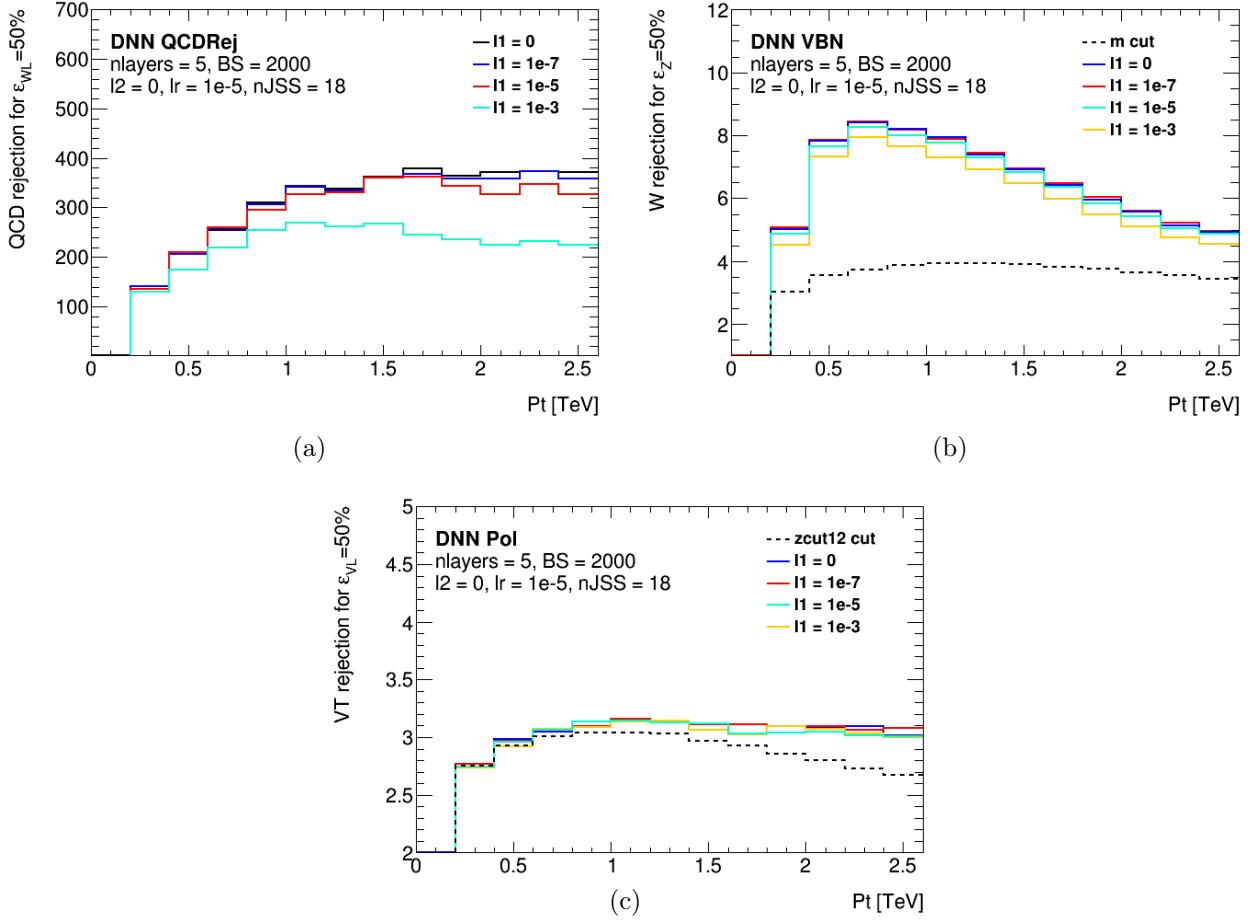


Figure 7.10: Background rejection as a function of transverse momentum for different $l1$ regularisation ($l1$) values for (a) DNN QCDRej tagger, (b) DNN VBN tagger and (c) DNN Pol tagger. $L1$ regularisation generally reduces the chances of overtraining; in this case it seems that not using any regularisation yields the best results for all three tagger types.

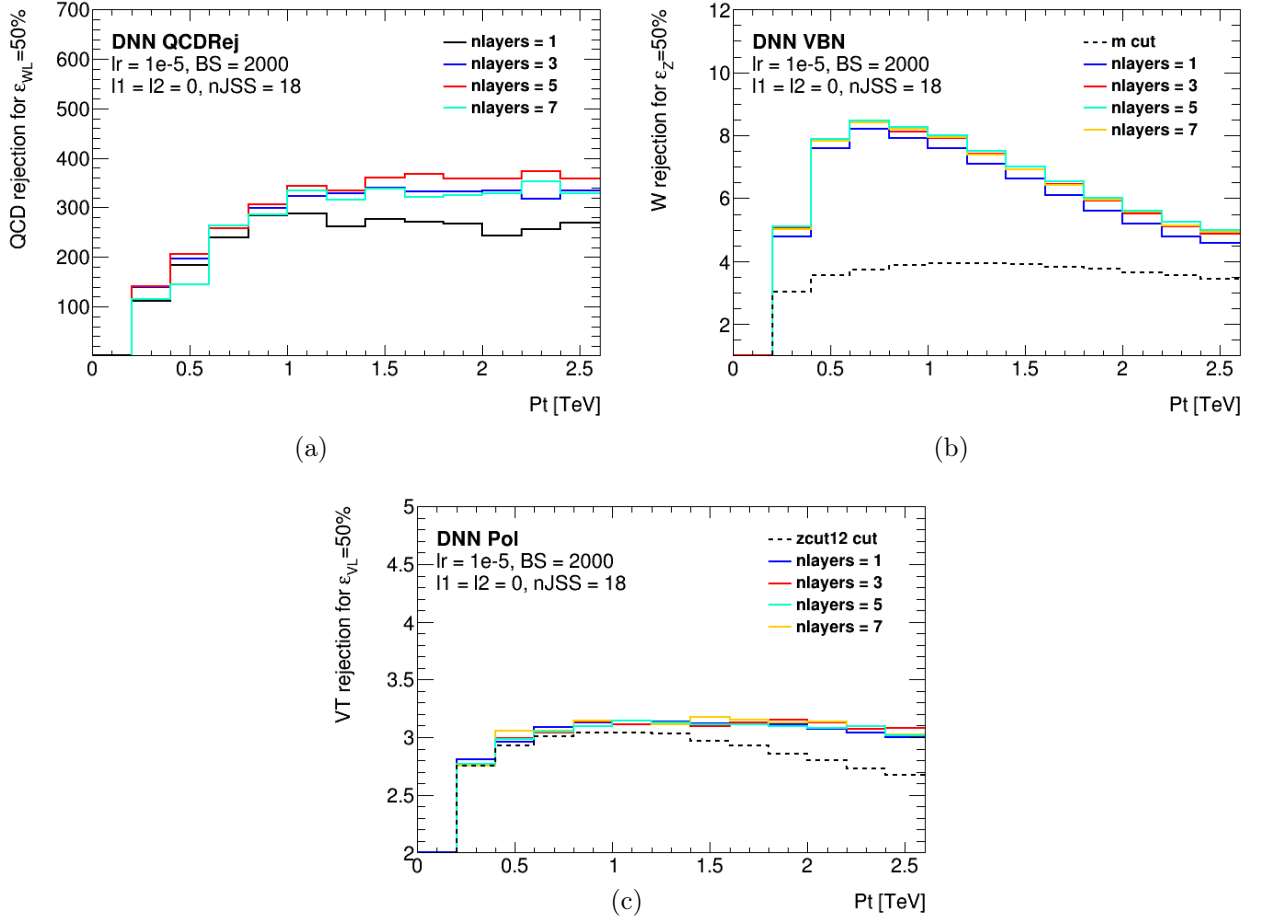


Figure 7.11: Background rejection as a function of transverse momentum for different architectures (number of layers, $nlayers$, and corresponding nodes, as detailed in table 7.3) for (a) DNN QCDRej tagger, (b) DNN VBN tagger and (c) DNN Pol tagger. A feed forward neural network with 5 hidden layers results in the best performance for all three taggers.

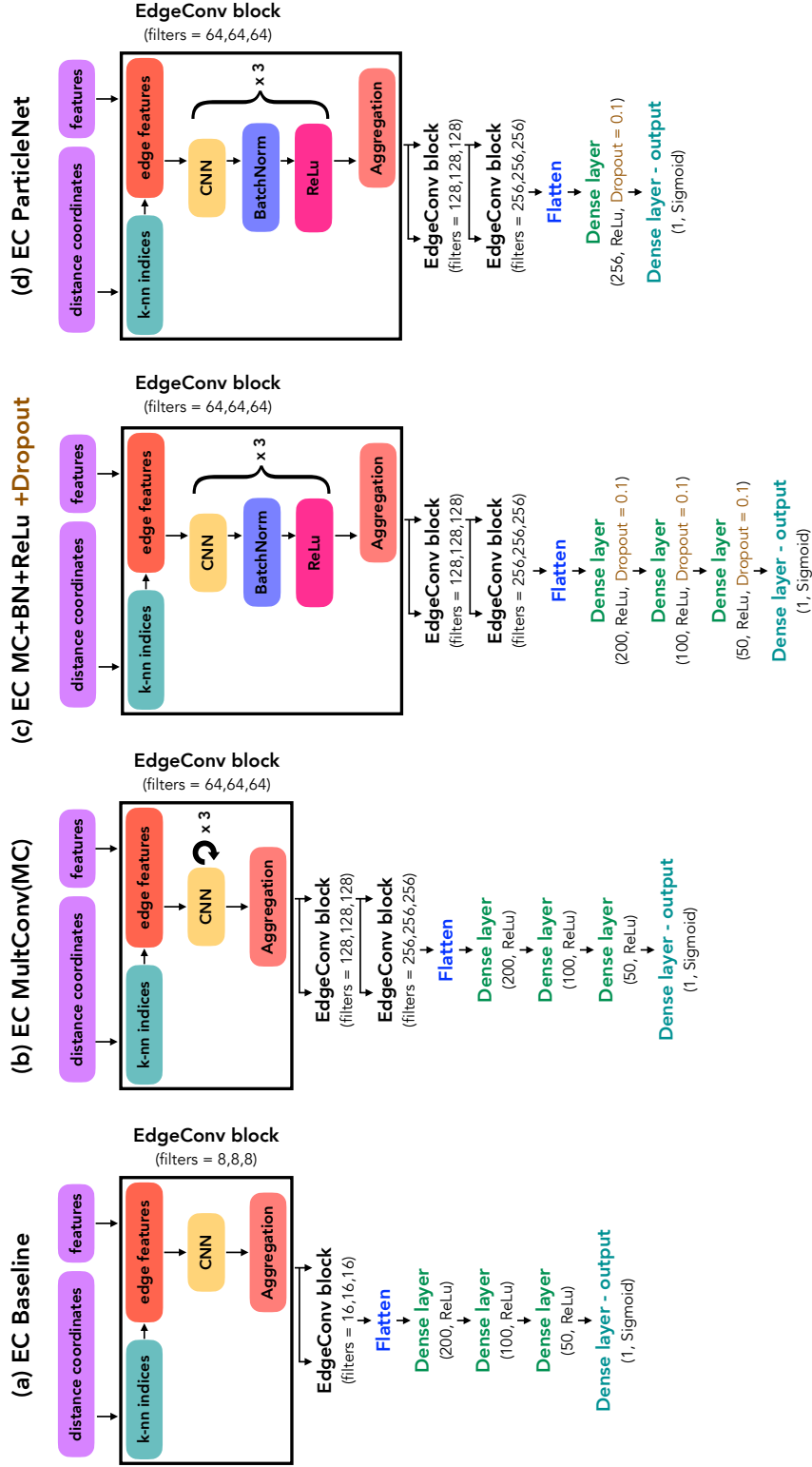


Figure 7.12: Different architectures tested: **EC Baseline** in (a), **EC MultConv(MC)** in (b), **EC MC+BN+ReLU** and **EC MC+BN+ReLU+Dropout** - differences between the two are marked in **brown** - in (c) and **EC ParticleNet** in (d).

| Jet constituent features | Definition |
|-----------------------------------|--|
| $\Delta\eta$ | Difference in pseudorapidity between the constituent and the jet axis |
| $\Delta\phi$ | Difference in azimuthal angle between the constituent and the jet axis |
| ΔR | Angular separation between the constituent and the jet axis $\sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ |
| $\log p_T$ | Logarithm of the constituent's p_T |
| $\log E^{const}$ | Logarithm of the constituent's energy |
| $\frac{\log p_T}{\log p_T^{jet}}$ | Logarithm of the constituent's p_T relative to the jet p_T |
| $\frac{\log E}{\log E^{jet}}$ | Logarithm of the constituent's energy relative to the jet energy |
| taste | Taste of the constituent as defined in section 3.3.1 |

Table 7.4: Description of the jet constituent features used as input for the training of the EdgeConv networks.

- The first strategy, and the one used throughout these studies, is to select a large number for $n_{Constituents}$, so that every jet in the sample is made up of a number of constituents $n_{Constituents}^{jet}$ smaller than the the constituent axis size; this is found to be $n_{Constituents} = 100$. Then, a number $n_{ghosts}^{jet} = n_{Constituents} - n_{Constituents}^{jet}$ of ghost particles⁽⁷⁾ randomly distributed within the jet radius ($R = 1.0$) is added as padding.
- The second strategy is similar to the first, the only difference is that instead of placing the ghost particles within $R = 1.0$ from the centre of the jet, the ghosts are placed outside of the jet's area. This strategy performs very poorly and was therefore discarded.

The number of features, $n_{Features}$, for each jet constituent is 8. In particular, other than the angular information that is needed to determine the nearest neighbours ($\Delta\eta_{const}$ and $\Delta\phi_{const}$), the remaining features are chosen based on studies performed in the aforementioned paper [113] and represent some of the main properties that characterise a constituent within a jet. The features and their definitions are summarised in table 7.4. It should be noted that, in addition to the features from the original paper, the UFO constituent's taste is also added to provide information on the origin of the constituent (coming from a charged particle⁽⁸⁾ or neutral particle).

The performance of the different configurations are evaluated in the same way used for the DNN tagger. In particular we find that the best architecture is the **EC MC+BN+Relu** (see figure 7.13)

⁽⁷⁾Ghost particles are particles with very small transverse momentum and energy so that they do not affect the reconstruction of jets. In this instance a value of $p_T = E = 2.7$ MeV was used.

⁽⁸⁾Where the taste is split between charged and combined depending on the presence or absence of energy sharing (see section 3.3.1)

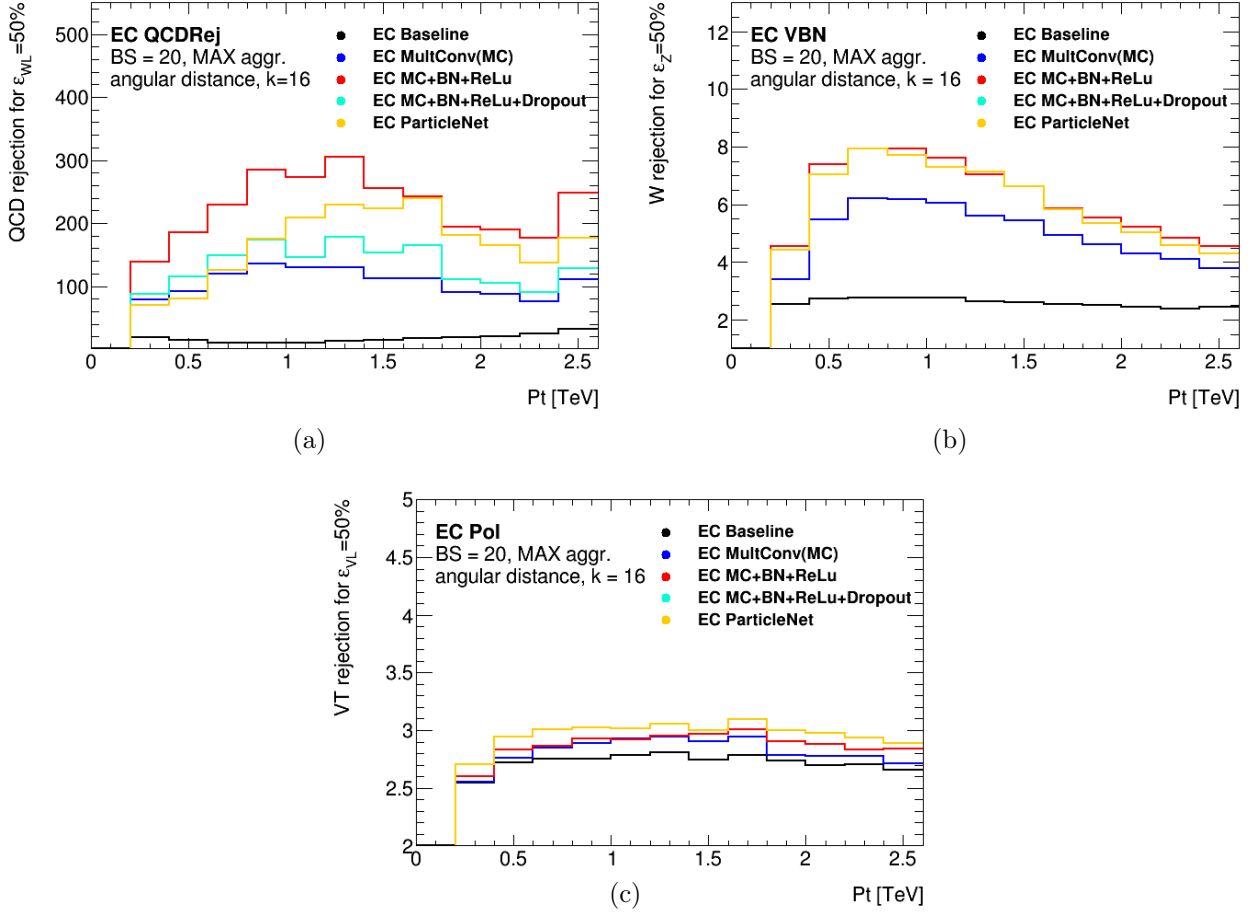


Figure 7.13: Background rejection as a function of transverse momentum for different EdgeConv architecture for (a) QCDRej tagger, (b) VBN tagger and (c) Pol tagger.

for all three tagger types and it is therefore the one used for the hyper-parameter optimisation and will be referred to as simply **EC** in the rest of the chapter.

The hyper-parameters tested are the following for all three tagger types: number of nearest neighbours, figure 7.14, batch size, figure 7.15, the distance metric (chosen amongst the usual euclidian distance, the k_T distance and the anti- k_T distance), figure 7.16, and the choice of aggregation function, figure 7.17.

The final architecture and hyper-parameter selection is, for all three classification tasks, the **EC MC+BN+ReLU** architecture with a batch size $BS=20$, a learning rate $lr=1e-5$, a choice of $k=16$ nearest neighbours found using the usual euclidian distance, and a choice of Max as the aggregation function.

7.5 Combined

To combine the results obtained in the previous two sections we create three new Combined Taggers (CT) which have the same architecture and hyper-parameters of the best DNNs found in section 7.3

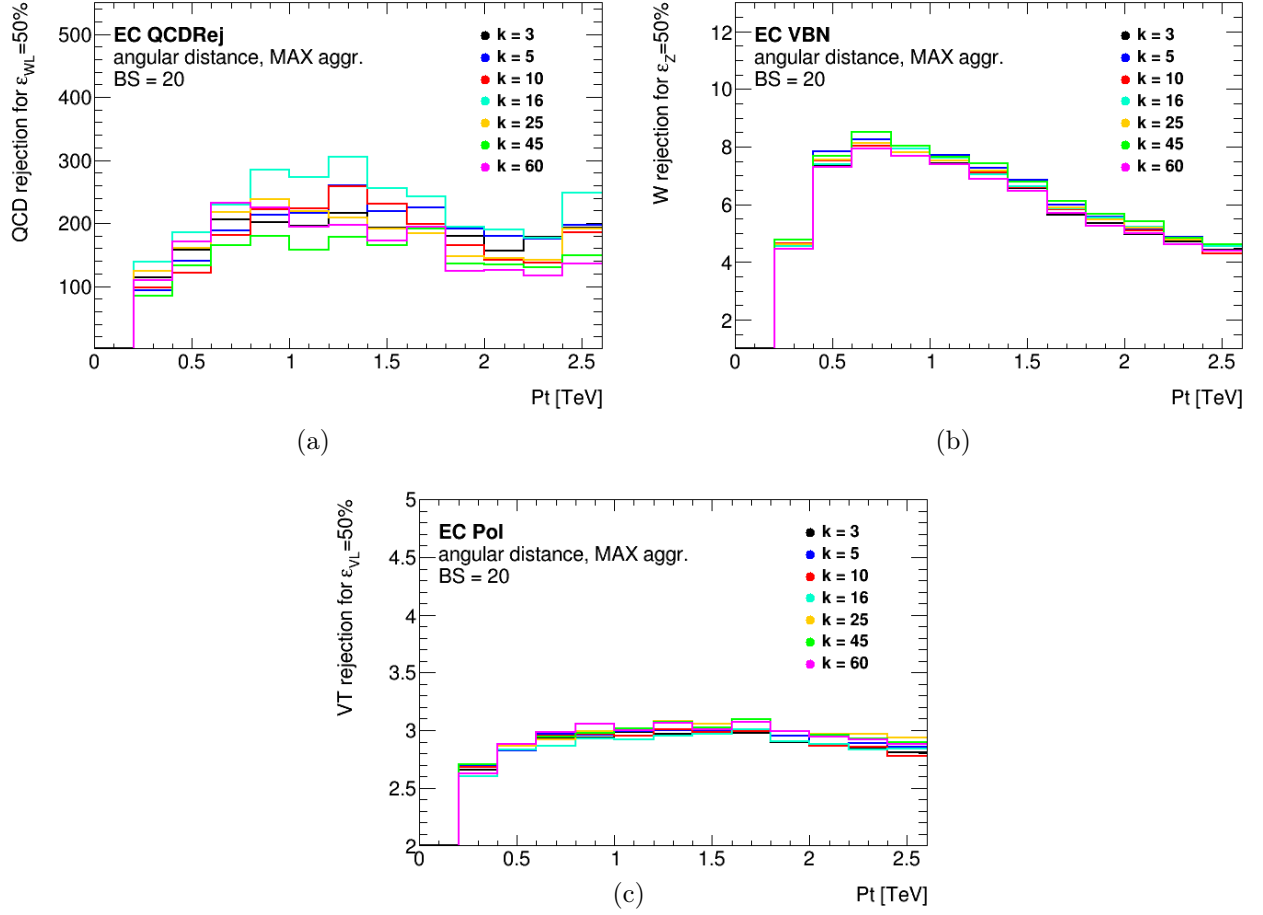


Figure 7.14: Background rejection as a function of transverse momentum for different numbers, k , of nearest neighbours for (a) EC QCDRej tagger, (b) EC VBN tagger and (c) EC Pol tagger. Best results are obtained with $k = 16$ for the QCDRej tagger whereas for the VBN and Pol tagger different choices of k yield similar results. As such $k = 16$ is chosen as the final value for all three tagger types.

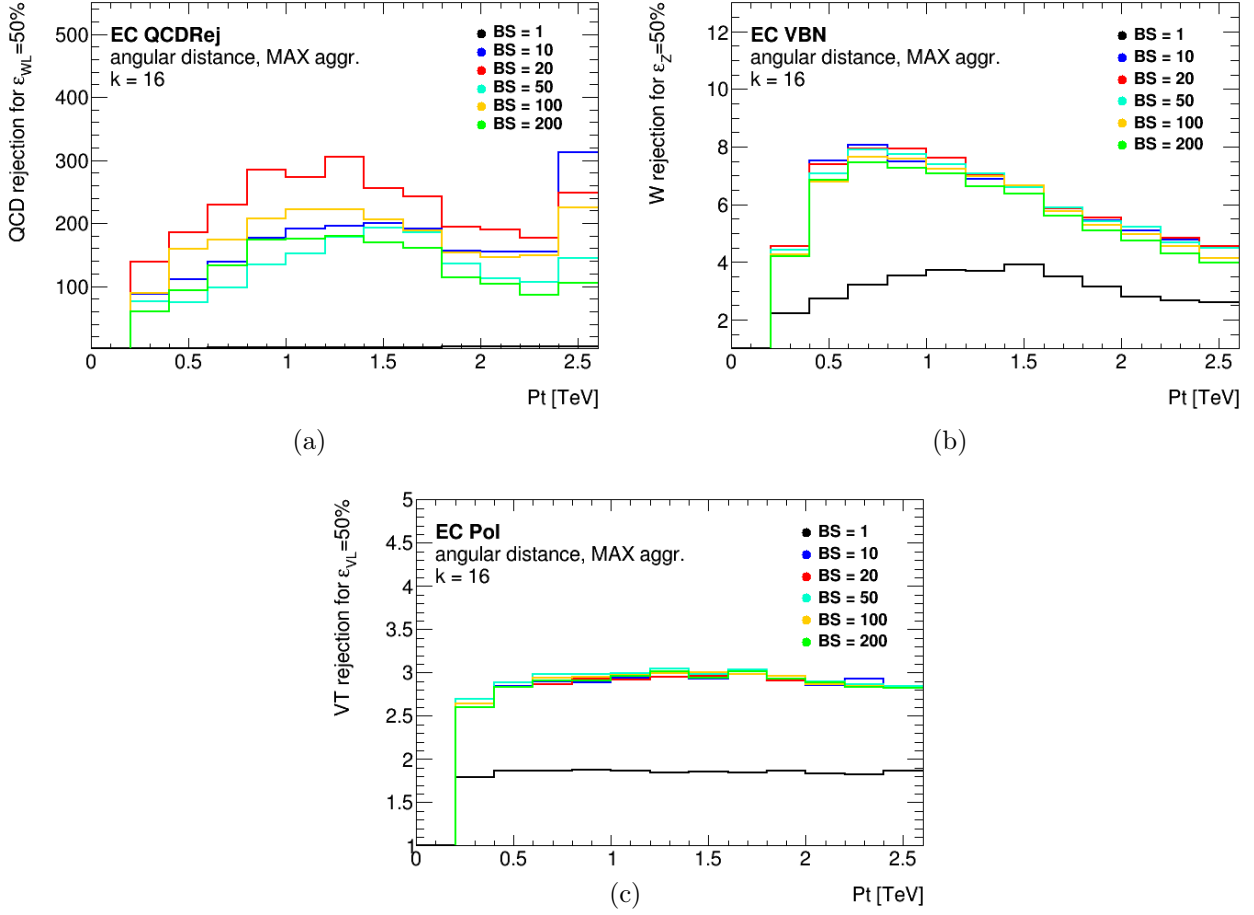


Figure 7.15: Background rejection as a function of transverse momentum for different batch size (BS) values for (a) EC QCDRej tagger, (b) EC VBN tagger and (c) EC Pol tagger. The results for batch size of 20 are the best or comparable for all three tagger types and is therefore chosen as the final architecture BS value.

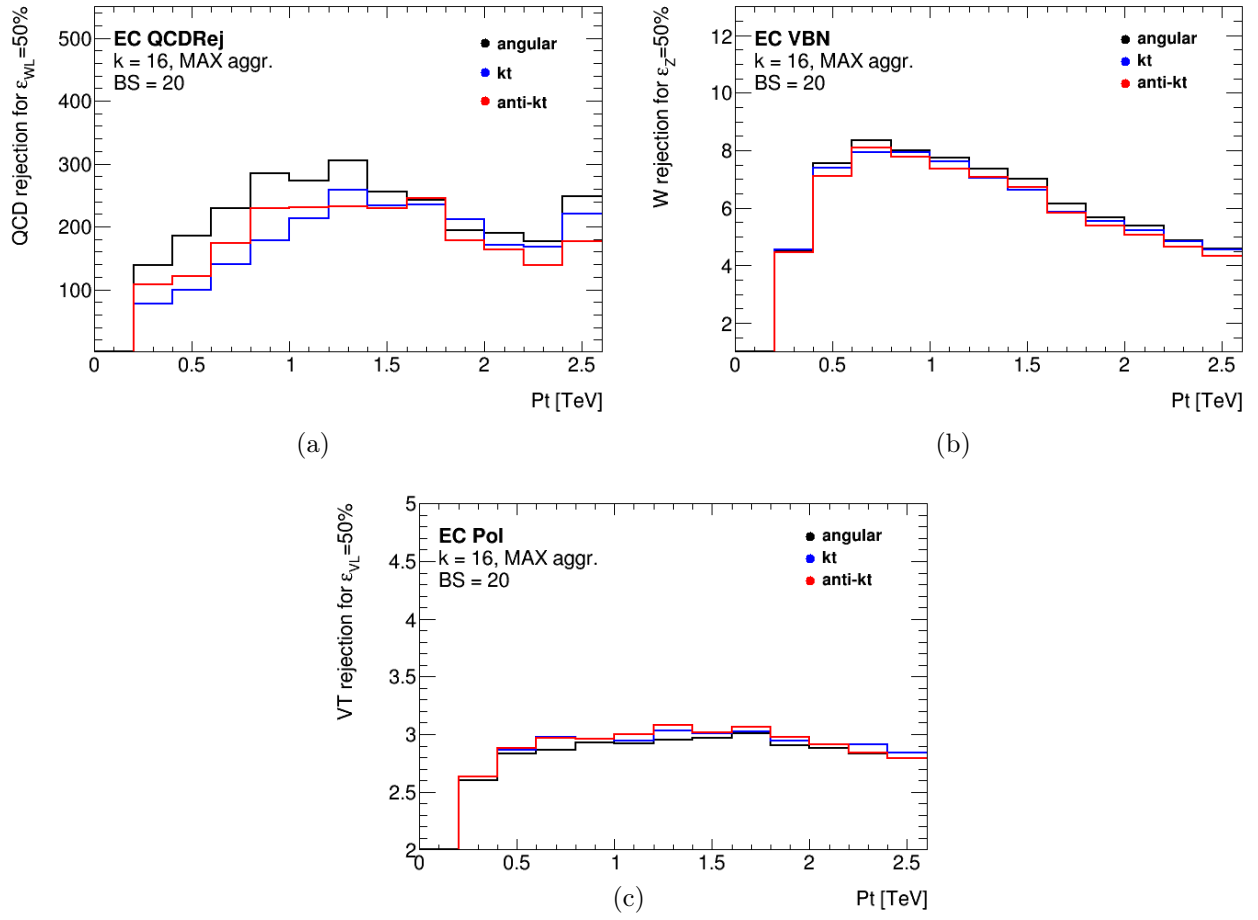


Figure 7.16: Background rejection as a function of transverse momentum for different choices of distance metric for (a) EC QCDRej tagger, (b) EC VBN tagger and (c) EC Pol tagger. The usual choice of euclidian distance is robust, less time intensive and yields the best or compatible results.

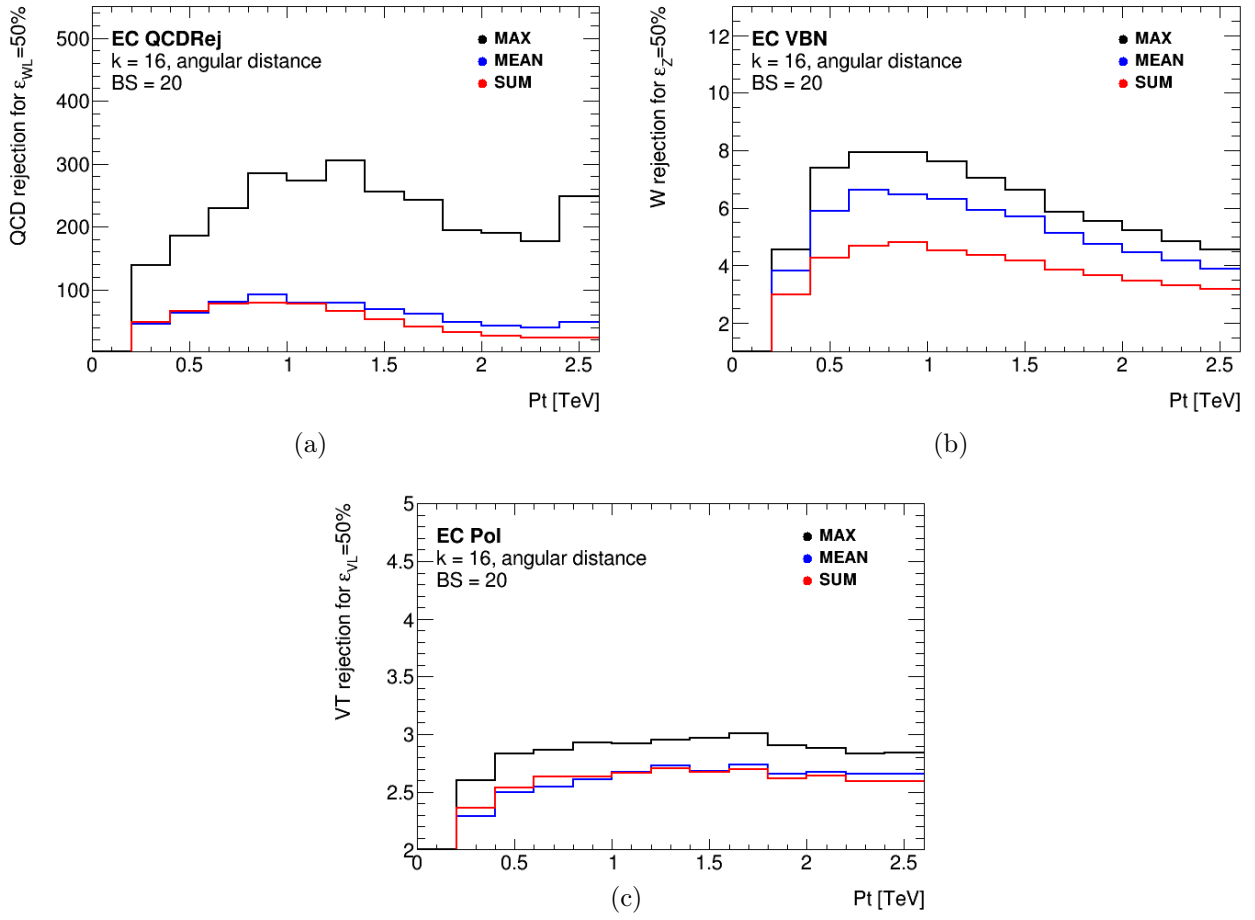


Figure 7.17: Background rejection as a function of transverse momentum for different choices of the aggregation function (Max, Mean and Sum) for (a) EC QCDRej tagger, (b) EC VBN tagger and (c) EC Pol tagger. The Max aggregation function yields the best results.

but with one or three more variables (on top of the original 18). The best EdgeConv models are used to predict the classification on the testing datasets and the resulting discriminant scores for each jet are used as additional inputs to the DNN. We test two configurations. In the first one, only the prediction of the same type of tagger is used ($N_{\text{pred}} = 1$): the prediction of the EC QCDRej model is used as an extra input to the CT QCDRej model, the prediction of the EC VBN model is used as an extra input to the CT VBN model, and so on. In the second configuration, all three predictions ($N_{\text{pred}}=3$), that is the measures of how consistent the jet is with a WL , with a Z and with a VL , are used as input variables to the three CT models.

The performance of the tagger is measured in terms of the background rejection for fixed (50%) signal efficiency, as a function of transverse momentum and can be seen in figure 7.18. The use of all three predictions yields a significant improvement for the QCDRej tagger: on top of high discriminant values for WL jets from the EC QCDRej tagger, most WL jets will also have higher scores from the EC VBN tagger and the EC Pol tagger with respect to a QCD jet, and this results in a reinforced prediction. The extra information brought by the different predictions doesn't result in an improvement of the performance of the CT VBN and CT Pol taggers, which are much more challenging and targeted tasks.

The final configuration for the Combined Taggers is therefore that of the DNN taggers with the addition of the predictions of the best three EdgeConv taggers.

A final comparison between the DNN taggers, EC taggers and CT taggers can be seen in figure 7.19. The background rejection as a function of p_T is also shown for the UFO and TCC cut-based taggers on three variables and for the best DNN taggers trained on TCC jets, which form the references to compare against. The Combined Tagger clearly outperforms all other configurations for the QCD rejection task with a four times improvement in comparison to the cut based taggers used in the previous VVJJ analysis described in chapter 5. The VBN and polarisation taggers are not as impacted by the additional tagging strategies but the discrimination power observed can still prove effective in the context of the analysis. The introduction of multiple variables however, also doubles the rejection power with respect to a single cut based tagger on the best variable available (see the ratios at the bottom of figures 7.19 of Pol and VBN tagger).

7.6 Outlook

The goal of tagging is to drastically reduce the background while minimally rejecting actual signal. In the context of the standard VVJJ analysis, this aspect is particularly relevant since it allows the mitigation of the 10^3 times larger cross section of QCD background with respect to the vector boson signal and thus make the observation of a resonant VV signal a possibility. Additional tagging abilities, such as VBN tagging and polarisation tagging, open the door to new measurements that extend the bump-hunt search performed in previous passes of the analysis. The goal of the studies shown above is to demonstrate the feasibility of the three classification tasks described.

But how do these results fit in the overall analysis strategy?

First, the difference in efficiency of the taggers between MC and real data needs to be tested in dedicated control regions. In particular, two real data samples can be used to highlight the ability

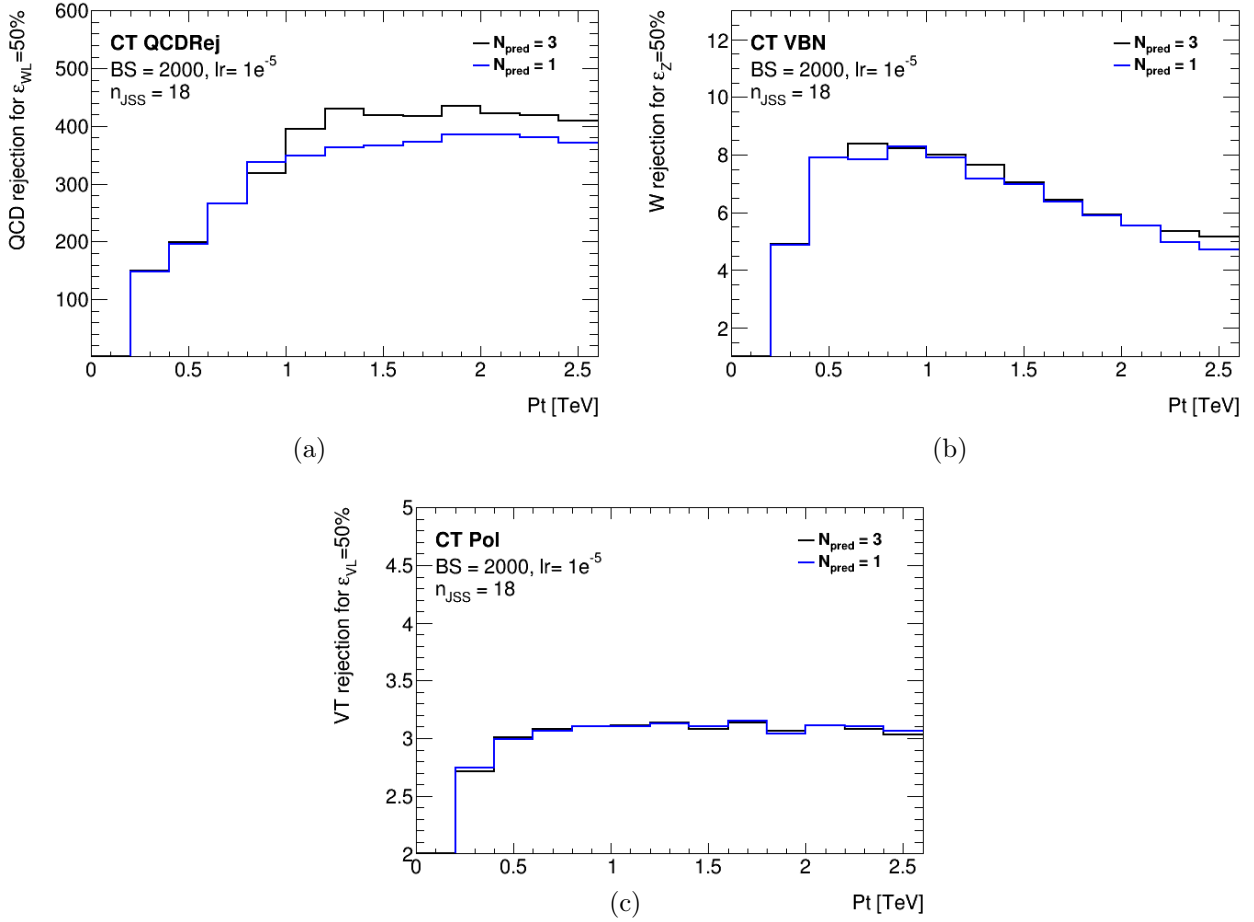


Figure 7.18: Background rejection as a function of transverse momentum for different numbers of predictions considered (N_{pred}) for (a) CT QCDRej tagger, (b) CT VBN tagger and (c) CT Pol tagger. Best results are obtained when using all three predictions for the QCDRej tagger whereas for the VBN and Pol taggers the use of one or three predictions doesn't make a significant difference. For consistency the final CT taggers all use the three predictions.

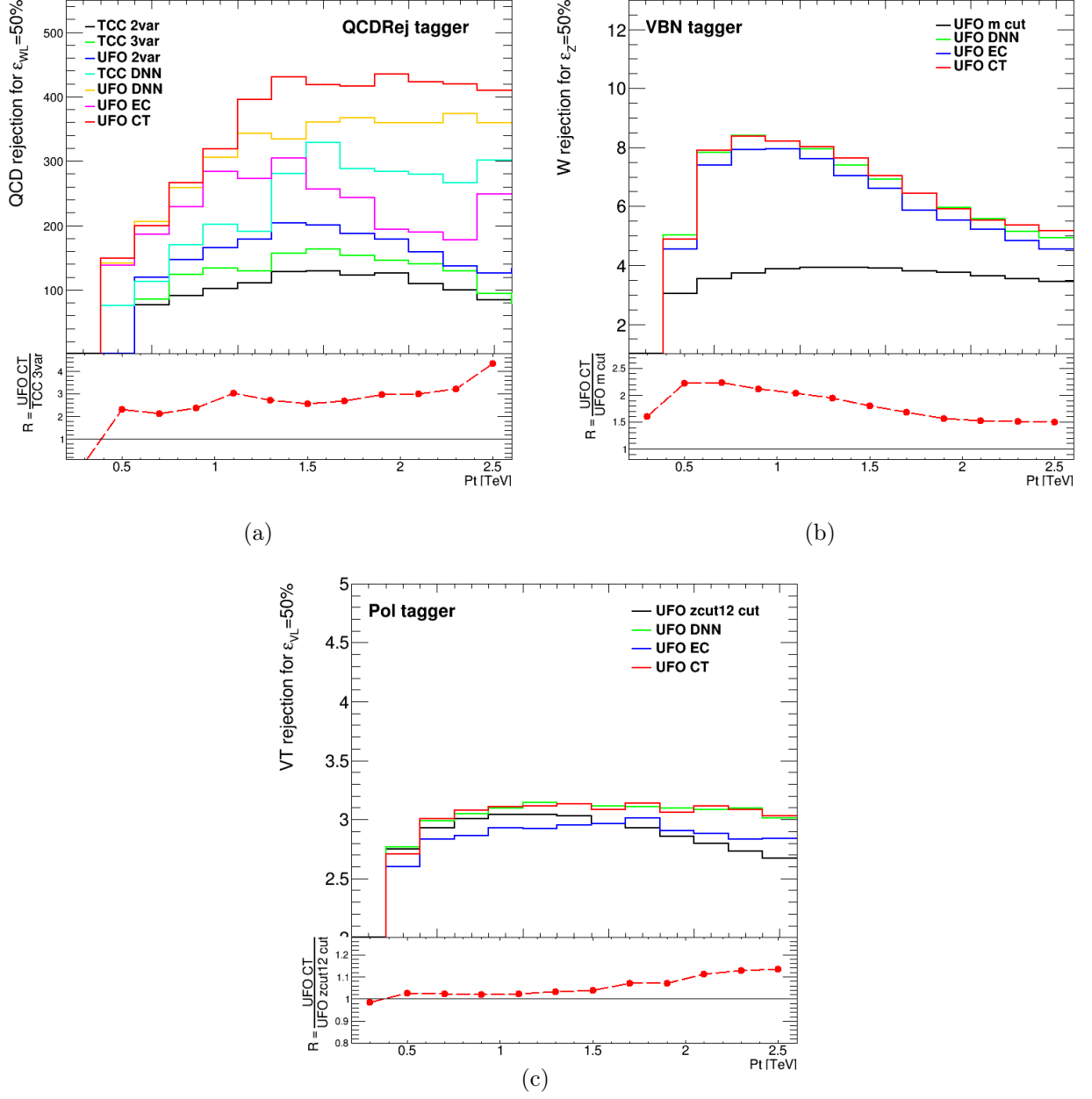


Figure 7.19: Background rejection as a function of transverse momentum for the best performing taggers developed with different algorithms in (a) for the CT QCDRej tagger, in (b) for the CT VBN tagger and in (c) for the CT Pol tagger. Best results are obtained using the Combined Tagger as can be seen from the ratio R plots.

of the taggers to correctly identify the three signal types: one control region sample enriched in final states with a vector boson plus jets (obtained following the procedure described in section 5.4) and in which the polarisation is primarily transverse [124], and one enriched in $t\bar{t}$ events, in which the polarisation is primarily longitudinal [125]. The V+jet sample allows the efficiency of the QCD rejection tagger⁽⁹⁾ and the VBN tagger to be tested for. For the VBN tagger a template method is used: templates of the mass distribution for W and Z bosons are extracted from MonteCarlo calculations. The tagger scale factors are then derived by fitting the tagged (corresponding to the Z bosons) and untagged (corresponding to the W bosons) data to the MonteCarlo templates. Finally, a comparison between the two samples (having opposite polarisation) gives information on the ability of the polarisation tagger to distinguish longitudinally polarised vector bosons from transversely polarised vector bosons. These tests on real data allow for the derivation of tagger scale factors to be used in the determination of the tagging efficiency uncertainty.

With the exception of the strategy modifications needed to account for new tagging possibilities, the rest of the analysis strategy follows the procedure used in the first pass of the VVJJ analysis (see chapter 5).

In conclusion, the development of two new tagger types and the dramatic improvement of the QCD rejection tagger pave the road for significant improvements and new goals in the context of a future iteration of the VVJJ analysis. In particular, since the analysis requires a double tag (one for each of the two bosons), any gain obtained is, in the end, squared. For the standard analysis, using the QCDRej tagger, this means an improvement of up to $4^2 = 16$ times. This leaves a large margin for the potential discovery of new physics.

⁽⁹⁾The method is described in detail in section 5.4

CONCLUSIONS

Di-boson interactions are a prime probe for searches of new physics beyond the Standard Model. One of the main motivations behind the choice of energy scale during the construction of the LHC was to test di-boson scattering processes and compare to SM expectations. The VVJJ analysis is one of the analyses concentrating on this type of interaction. In particular, it looks for a resonance in the dijet mass of pairs of tagged, fully hadronically decaying, vector bosons.

The first pass of the VVJJ analysis over Run2 data didn't reveal deviations from the Standard Model of particle physics. The introduction of a different strategy, however, resulted in dramatic improvements in the extent of exclusion limits for a number of theoretical models. In particular, the creation of Track-CaloClusters and the development of a three variable cut-based tagger were the changes that resulted in the largest gain. Indeed both the use of TCC jets and the use of an additional variables in the tagger algorithm resulted in a dramatic increase of the tagger ability to select signal events while rejecting background events.

The development of United Flow Objects motivates a future iteration of the VVJJ analysis on the full Run2 or future datasets. On top of using these new objects, the analysis revisited its strategy for the search to be sensitive to the polarisation and flavour of the vector bosons involved. To achieve tagging on these properties, in addition to the usual cut-based QCD rejection taggers, several Machine Learning algorithms were tested. The goal of this thesis is to demonstrate a proof of concept for these new taggers.

From the introduction of UFO jets and the use of Machine Learning taggers, we observe a significant improvement in the ability to reject QCD background (up to four times better). The task of identifying the origin of the jet (W or Z) and its polarisation state proved more challenging but the new strategies still resulted in a two fold improvement for the former and up to 15% improvement for the latter. These polarisation studies, in particular, are the first attempt to identify the polarisation of fully hadronic decays of vector bosons, thus marking an important benchmark in the development of analyses probing the precise expectations in the amplitudes of polarised vector boson interactions.

ACKNOWLEDGEMENTS

The work presented in this thesis is the result of four beautiful and intense years. It is a labour of love (mine) and patience (of the incredible people that surround me) and is my proudest achievement to date.

No man is an island and I wouldn't be here today without the people that have supported and guided me during this time.

First, thanks to professor Simonetta Gentile without whom I would have never entered the field of high energy physics and for first introducing me to the ATLAS group of the Université de Genève.

Next, I wish to thank professor Giuseppe Iacobucci for welcoming me to the team and for giving me the opportunity of doing a PhD under his supervision: these four years have been the most formative of my life so far. A big thank you to the entire ATLAS group at UniGe for their patience and for answering all of the silly questions I've asked. In particular, thank you to Roland and Johnny for their guidance during this adventure, they have taught me a lot.

Very special thanks to my supervisor and mentor, Steven Schramm without whose vision this work wouldn't exist and without whom I would have probably never considered enrolling in a doctorate program. I cannot say how much I have learned from him during this time. I couldn't have met anyone better than him to teach me about pretty much any physics/computing topic and countless others. In particular, his work ethic, passion and determination in doing what is right will be a compass for me in all my future endeavours.

Thanks to Riccardo, Marco and Juno for the laughs, the coffees/teas, and the support, both in person and remotely (thank you COVID pandemic). It wouldn't have been this fun without you.

Thanks to Charlotte, Giulia and Letizia for being my trusted job search support squad/ motivation letter review committee/ interview prepping team. Thanks to Rebecka for always making me laugh no matter the mood I'm in. Thanks to Corentin, Romane and Stefano for being amazing friends that ALSO climb (could you ask for any better?) and Heather (even though she doesn't climb). Thanks to Beatrice, Paolo, Guido, Andrea C., and Briano for bringing a little (crazy) piece of Italy here in Switzerland. Thanks to Costanza and Matteo: you abandoned the swiss ship but brought Gregorio into our lives so you're forgiven. All you guys have made Geneva feel like home.

Thanks to Prisca, Margherita, Valentine and Alessandro for always being there when I needed you, whether it was for a good laugh, an amazing trip or just spending time together (I know it's the same acknowledgement as four years ago but nothing has changed since and I hope nothing ever will).

Thanks to my family (Mamma, Papà, Franci, Gianlu, Nonna et Nonno) for their unconditional love and their unlimited amount of patience: you are the pillars of my existence.

Finally, I have to thank Andrea. For everything.

Once again, thanks. I wouldn't be here without all of you.

Bibliography

- [1] ATLAS Collaboration. *Identification of boosted, hadronically-decaying W and Z bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-033>. Geneva: CERN, Aug. 2015. URL: <https://cds.cern.ch/record/2041461>.
- [2] ATLAS Collaboration. *Boosted hadronic top identification at ATLAS for early 13 TeV data*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-053>. Geneva: CERN, Dec. 2015. URL: <https://cds.cern.ch/record/2116351>.
- [3] ATLAS Collaboration. *Quark versus Gluon Jet Tagging Using Charged Particle Multiplicity with the ATLAS Detector*. Tech. rep. ATL-PHYS-PUB-2017-009. Geneva: CERN, May 2017. URL: <https://cds.cern.ch/record/2263679>.
- [4] S.L. Glashow. “Partial Symmetries of Weak Interactions”. In: *Nucl. Phys.* 22 (1961), p. 579. DOI: [10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [5] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (1967), p. 1264. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [6] Abdus Salam. “Weak and Electromagnetic Interactions”. In: *Conf. Proc.* C680519 (1968), pp. 367–377.
- [7] ALEPH Collaboration. “Precision electroweak measurements on the Z resonance”. In: *Physics Reports* 427.5-6 (May 2006), 257–454. ISSN: 0370-1573. DOI: [10.1016/j.physrep.2005.12.006](https://doi.org/10.1016/j.physrep.2005.12.006). URL: <http://dx.doi.org/10.1016/j.physrep.2005.12.006>.
- [8] *Standard Model Image*. URL: https://en.wikipedia.org/wiki/Standard_Model.
- [9] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Physical Review Letters* 81.8 (Aug. 1998), 1562–1567. ISSN: 1079-7114. DOI: [10.1103/physrevlett.81.1562](https://doi.org/10.1103/physrevlett.81.1562). URL: <http://dx.doi.org/10.1103/PhysRevLett.81.1562>.
- [10] Riccardo Poggi. “Top-quark pair production cross-section measurements in the all-hadronic decay channel at the ATLAS experiment and hardware-based track reconstruction for the ATLAS trigger HL-LHC upgrade.” PhD thesis. Université de Genève, 2020. URL: <http://archive-ouverte.unige.ch/unige:151144>.
- [11] J. Beringer et al. “Quantum Chromodynamics”. In: *Chin. Phys.* C38 (2014), p. 090001. DOI: [10.1088/1674-1137/38/9/090001](https://doi.org/10.1088/1674-1137/38/9/090001).
- [12] Antonio Pich. “The Standard Model of Electroweak Interactions; rev. version”. In: (Feb. 2005), 48 p. DOI: [10.5170/CERN-2006-003.1](https://doi.org/10.5170/CERN-2006-003.1). URL: <https://cds.cern.ch/record/819632>.
- [13] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).

- [14] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [15] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: [10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.
- [16] Makoto Kobayashi and Toshihide Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: [10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [17] Murray Gell-Mann, Pierre Ramond, and Richard Slansky. *Complex Spinors and Unified Theories*. 2013. arXiv: [1306.4669](https://arxiv.org/abs/1306.4669) [hep-th].
- [18] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. “Matter and antimatter in the universe”. In: *New Journal of Physics* 14.9 (Sept. 2012), p. 095012. ISSN: 1367-2630. DOI: [10.1088/1367-2630/14/9/095012](https://doi.org/10.1088/1367-2630/14/9/095012). URL: <http://dx.doi.org/10.1088/1367-2630/14/9/095012>.
- [19] STEPHEN P. MARTIN. “A SUPERSYMMETRY PRIMER”. In: *Advanced Series on Directions in High Energy Physics* (July 1998), 1?98. ISSN: 1793-1339. DOI: [10.1142/9789812839657_0001](https://doi.org/10.1142/9789812839657_0001). URL: http://dx.doi.org/10.1142/9789812839657_0001.
- [20] H. Georgi and S. L. Glashow. “Unity of All Elementary Particle Forces”. In: *Phys. Rev. Lett.* 32 (1974), pp. 438–441. DOI: [10.1103/PhysRevLett.32.438](https://doi.org/10.1103/PhysRevLett.32.438).
- [21] Harald Fritzsch and Peter Minkowski. “Unified Interactions of Leptons and Hadrons”. In: *Annals Phys.* 93 (1975), pp. 193–266. DOI: [10.1016/0003-4916\(75\)90211-0](https://doi.org/10.1016/0003-4916(75)90211-0).
- [22] Chris Malena Delitzsch. “Search for Diboson Resonances in the Fully Hadronic Final State Using Jet Substructure Techniques in 8 and 13 TeV Proton-Proton Collisions with the ATLAS Detector”. Presented 02 Dec 2016. PhD thesis. Université de Genève, 2016. URL: <http://cds.cern.ch/record/2242565>.
- [23] Simon Catterall et al. *MCRG Minimal Walking Technicolor*. 2010. arXiv: [1010.5909](https://arxiv.org/abs/1010.5909) [hep-ph].
- [24] Estia Eichten and Kenneth Lane. “Low-scale technicolor at the Tevatron and LHC”. In: *Physics Letters B* 669.3-4 (Nov. 2008), 235?238. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2008.09.047](https://doi.org/10.1016/j.physletb.2008.09.047). URL: <http://dx.doi.org/10.1016/j.physletb.2008.09.047>.
- [25] J. R. Andersen et al. “Discovering Technicolor”. In: *The European Physical Journal Plus* 126.9 (Sept. 2011), p. 48. ISSN: 2190-5444. DOI: [10.1140/epjp/i2011-11081-1](https://doi.org/10.1140/epjp/i2011-11081-1). URL: <http://dx.doi.org/10.1140/epjp/i2011-11081-1>.
- [26] Lisa Randall and Raman Sundrum. “Large Mass Hierarchy from a Small Extra Dimension”. In: *Physical Review Letters* 83.17 (Oct. 1999), 3370?3373. ISSN: 1079-7114. DOI: [10.1103/physrevlett.83.3370](https://doi.org/10.1103/physrevlett.83.3370). URL: <http://dx.doi.org/10.1103/PhysRevLett.83.3370>.
- [27] Lisa Randall and Raman Sundrum. “An Alternative to Compactification”. In: *Physical Review Letters* 83.23 (Dec. 1999), 4690?4693. ISSN: 1079-7114. DOI: [10.1103/physrevlett.83.4690](https://doi.org/10.1103/physrevlett.83.4690). URL: <http://dx.doi.org/10.1103/PhysRevLett.83.4690>.

- [28] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo. “Experimental probes of localized gravity: On and off the wall”. In: *Physical Review D* 63.7 (Mar. 2001), p. 48. ISSN: 1089-4918. DOI: [10.1103/PhysRevD.63.075004](https://doi.org/10.1103/PhysRevD.63.075004). URL: <http://dx.doi.org/10.1103/PhysRevD.63.075004>.
- [29] ATLAS Collaboration. “Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2015.12 (Dec. 2015), 1?39. ISSN: 1029-8479. DOI: [10.1007/jhep12\(2015\)055](https://doi.org/10.1007/jhep12(2015)055). URL: [http://dx.doi.org/10.1007/JHEP12\(2015\)055](http://dx.doi.org/10.1007/JHEP12(2015)055).
- [30] ATLAS Collaboration. *Search for resonances with boson-tagged jets in 15.5 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2016-055>. Geneva: CERN, Aug. 2016. URL: <https://cds.cern.ch/record/2206137>.
- [31] ATLAS Collaboration. “Search for diboson resonances with boson-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Physics Letters B* 777 (Feb. 2018), 91?113. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2017.12.011](https://doi.org/10.1016/j.physletb.2017.12.011). URL: <http://dx.doi.org/10.1016/j.physletb.2017.12.011>.
- [32] ATLAS Collaboration. “Search for diboson resonances in hadronic final states in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2019.9 (Sept. 2019), 1?39. ISSN: 1029-8479. DOI: [10.1007/jhep09\(2019\)091](https://doi.org/10.1007/jhep09(2019)091). URL: [http://dx.doi.org/10.1007/JHEP09\(2019\)091](http://dx.doi.org/10.1007/JHEP09(2019)091).
- [33] G. Altarelli, B. Mele, and M. Ruiz-Altaba. “Searching for new heavy vector bosons in $p\bar{p}$ colliders”. In: *Zeitschrift für Physik C Particles and Fields* 45.1 (Mar. 1989), pp. 109–121. ISSN: 1431-5858. DOI: [10.1007/BF01556677](https://doi.org/10.1007/BF01556677). URL: <https://doi.org/10.1007/BF01556677>.
- [34] E. Eichten et al. “Supercollider physics”. In: *Rev. Mod. Phys.* 56 (4 Oct. 1984), pp. 579–707. DOI: [10.1103/RevModPhys.56.579](https://doi.org/10.1103/RevModPhys.56.579). URL: <https://link.aps.org/doi/10.1103/RevModPhys.56.579>.
- [35] Walter D. Goldberger and Mark B. Wise. “Modulus Stabilization with Bulk Fields”. In: *Physical Review Letters* 83.24 (Dec. 1999), 4922?4925. ISSN: 1079-7114. DOI: [10.1103/PhysRevLett.83.4922](https://doi.org/10.1103/PhysRevLett.83.4922). URL: <http://dx.doi.org/10.1103/PhysRevLett.83.4922>.
- [36] Walter D. Goldberger and Mark B. Wise. “Phenomenology of a stabilized modulus”. In: *Physics Letters B* 475.3-4 (Mar. 2000), 275?279. ISSN: 0370-2693. DOI: [10.1016/S0370-2693\(00\)00099-X](https://doi.org/10.1016/S0370-2693(00)00099-X). URL: [http://dx.doi.org/10.1016/S0370-2693\(00\)00099-X](http://dx.doi.org/10.1016/S0370-2693(00)00099-X).
- [37] Tao Han, Joseph D. Lykken, and Ren-Jie Zhang. “On Kaluza-Klein states from large extra dimensions”. In: *Phys. Rev. D* 59 (1999), p. 105006. DOI: [10.1103/PhysRevD.59.105006](https://doi.org/10.1103/PhysRevD.59.105006). arXiv: [hep-ph/9811350](https://arxiv.org/abs/hep-ph/9811350) [hep-ph].
- [38] Duccio Pappadopulo et al. “Heavy Vector Triplets: Bridging Theory and Data”. In: *JHEP* 09 (2014), p. 060. DOI: [10.1007/JHEP09\(2014\)060](https://doi.org/10.1007/JHEP09(2014)060). arXiv: [1402.4431](https://arxiv.org/abs/1402.4431) [hep-ph].
- [39] O. Buning et al. “LHC Design Report. 1. The LHC Main Ring”. In: (Nov. 2004). Ed. by O. Buning et al. DOI: [10.5170/CERN-2004-003-V-2](https://doi.org/10.5170/CERN-2004-003-V-2).

- [40] O. Buning et al. “LHC Design Report. 2. The LHC infrastructure and general services”. In: (Nov. 2004). Ed. by O. Buning et al. DOI: [10.5170/CERN-2004-003-V-2](https://doi.org/10.5170/CERN-2004-003-V-2).
- [41] M. Benedikt et al. “LHC Design Report. 3. The LHC injector chain”. In: (Dec. 2004). Ed. by M. Benedikt et al. DOI: [10.5170/CERN-2004-003-V-3](https://doi.org/10.5170/CERN-2004-003-V-3).
- [42] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [43] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [44] K. Aamodt et al. “The ALICE experiment at the CERN LHC”. In: *JINST* 3 (2008), S08002. DOI: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [45] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008), S08005. DOI: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [46] CERN. *Map of the CERN accelerator complex*. URL: <http://public-archive.web.cern.ch/public-archive/en/research/AccelComplex-en.html>.
- [47] ATLAS Collaboration. *Luminosity Run1*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- [48] ATLAS Collaboration. *Luminosity Run2*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [49] HiLumi. *Time Scale*. URL: <https://project-hl-lhc-industry.web.cern.ch/content/project-schedule>.
- [50] ATLAS Collaboration. *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS. Geneva: CERN, 1997. URL: <https://cds.cern.ch/record/331063>.
- [51] Silvia Masciocchi. *Semiconductor detector*. URL: https://www.physi.uni-heidelberg.de/~sma/teaching/GraduateDays2017/sma_Detectors_3_Semiconductors.pdf.
- [52] G Schlager. “The Energy Response of the ATLAS Calorimeter System”. Presented on 04 Dec 2006. PhD thesis. Vienna University of Technology, 2006. URL: <http://cds.cern.ch/record/998248>.
- [53] ATLAS Collaboration. “ATLAS calorimeter performance Technical Design Report”. In: (Dec. 1996).
- [54] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *The European Physical Journal C* 77.7 (July 2017), S08005. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5004-5](https://doi.org/10.1140/epjc/s10052-017-5004-5). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5004-5>.
- [55] ATLAS Collaboration. “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run2”. In: *Eur. Phys. J. C* 77.10 (2017), p. 673. DOI: [10.1140/epjc/s10052-017-5225-7](https://doi.org/10.1140/epjc/s10052-017-5225-7). arXiv: [1704.07983](https://arxiv.org/abs/1704.07983) [hep-ex].

- [56] Azriel Rosenfeld and John L. Pfaltz. “Sequential Operations in Digital Picture Processing”. In: *J. ACM* 13.4 (Oct. 1966), pp. 471–494. ISSN: 0004-5411. DOI: [10.1145/321356.321357](https://doi.acm.org/10.1145/321356.321357). URL: <http://doi.acm.org/10.1145/321356.321357>.
- [57] ATLAS Collaboration. *Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at $\sqrt{s} = 7$ TeV*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2010-072>. Geneva: CERN, July 2010. URL: <https://cds.cern.ch/record/1281363>.
- [58] Nazar Bartosik. *Anatomy of a b jet*. URL: <https://it.wikipedia.org/wiki/B-tagging>.
- [59] ATLAS Collaboration. *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjects in ATLAS*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2020-019>. Geneva: CERN, July 2020. URL: <https://cds.cern.ch/record/2724739>.
- [60] ATLAS Collaboration. *Simulation-based extrapolation of b -tagging calibrations towards high transverse momenta in the ATLAS experiment*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-003>. Geneva: CERN, Mar. 2021. URL: <http://cds.cern.ch/record/2753444>.
- [61] Sofia Adorni. *Increasing track reconstruction efficiency in dense environments at ATLAS*. Tech. rep. Geneva: CERN, Apr. 2019. URL: <https://cds.cern.ch/record/2669951>.
- [62] ATLAS Collaboration. *Improving jet substructure performance in ATLAS using Track-CaloClusters*. Tech. rep. ATL-PHYS-PUB-2017-015. Geneva: CERN, July 2017. URL: <https://cds.cern.ch/record/2275636>.
- [63] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77.7 (2017), p. 466. DOI: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2). arXiv: [1703.10485](https://arxiv.org/abs/1703.10485) [hep-ex].
- [64] ATLAS Collaboration. “Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions in the ATLAS detector”. In: *The European Physical Journal C* 77.4 (Apr. 2017), p. 673. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-4780-2](https://doi.org/10.1140/epjc/s10052-017-4780-2). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-4780-2>.
- [65] ATLAS Collaboration. “Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton-proton collisions”. In: *Eur. Phys. J. C* 81 (Sept. 2020), 334. 47 p. DOI: [10.1140/epjc/s10052-021-09054-3](https://doi.org/10.1140/epjc/s10052-021-09054-3). arXiv: [2009.04986](https://arxiv.org/abs/2009.04986). URL: <http://cds.cern.ch/record/2730141>.
- [66] 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. DOI: [10.1007/JHEP09\(2013\)076](https://doi.org/10.1007/JHEP09(2013)076). arXiv: [1306.4945](https://arxiv.org/abs/1306.4945) [hep-ex].
- [67] Gavin P. Salam. “Towards Jetography”. In: *Eur. Phys. J. C* 67 (2010), pp. 637–686. DOI: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6). arXiv: [0906.1833](https://arxiv.org/abs/0906.1833) [hep-ph].

- [68] John E. Huth et al. “Toward a standardization of jet definitions”. In: *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade, Snowmass, Colorado, June 25-July 13, 1990*, FERMILAB-CONF-90-249-E, FNAL-C-90-249-E (), pp. 134–136. URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-249.
- [69] M. Wobisch and T. Wengler. “Hadronization corrections to jet cross-sections in deep inelastic scattering”. In: *Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)*. Apr. 1998. arXiv: [hep-ph/9907280](https://arxiv.org/abs/hep-ph/9907280).
- [70] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Anti- k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). arXiv: [0802.1189](https://arxiv.org/abs/0802.1189) [[hep-ph](#)].
- [71] Yuri L. Dokshitzer et al. “Better jet clustering algorithms”. In: *JHEP* 08 (1997), p. 001. DOI: [10.1088/1126-6708/1997/08/001](https://doi.org/10.1088/1126-6708/1997/08/001). arXiv: [hep-ph/9707323](https://arxiv.org/abs/hep-ph/9707323) [[hep-ph](#)].
- [72] ATLAS Collaboration. “Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector”. In: *The European Physical Journal C* 76.6 (June 2016), pp. 134–136. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-016-4126-5](https://doi.org/10.1140/epjc/s10052-016-4126-5). URL: <http://dx.doi.org/10.1140/epjc/s10052-016-4126-5>.
- [73] David Krohn, Jesse Thaler, and Lian-Tao Wang. “Jet Trimming”. In: *JHEP* 02 (2010), p. 084. DOI: [10.1007/JHEP02\(2010\)084](https://doi.org/10.1007/JHEP02(2010)084). arXiv: [0912.1342](https://arxiv.org/abs/0912.1342) [[hep-ph](#)].
- [74] Andrew J. Larkoski et al. “Soft drop”. In: *Journal of High Energy Physics* 2014.5 (May 2014), p. 005. ISSN: 1029-8479. DOI: [10.1007/jhep05\(2014\)146](https://doi.org/10.1007/jhep05(2014)146). URL: [http://dx.doi.org/10.1007/JHEP05\(2014\)146](http://dx.doi.org/10.1007/JHEP05(2014)146).
- [75] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. “Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches”. In: *Phys. Rev. D* 81 (2010), p. 094023. DOI: [10.1103/PhysRevD.81.094023](https://doi.org/10.1103/PhysRevD.81.094023). arXiv: [0912.0033](https://arxiv.org/abs/0912.0033) [[hep-ph](#)].
- [76] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The catchment area of jets”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), 005?005. ISSN: 1029-8479. DOI: [10.1088/1126-6708/2008/04/005](https://doi.org/10.1088/1126-6708/2008/04/005). URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/005>.
- [77] Peter Berta et al. “Particle-level pileup subtraction for jets and jet shapes”. In: *Journal of High Energy Physics* 2014.6 (June 2014), p. 005. ISSN: 1029-8479. DOI: [10.1007/jhep06\(2014\)092](https://doi.org/10.1007/jhep06(2014)092). URL: [http://dx.doi.org/10.1007/JHEP06\(2014\)092](http://dx.doi.org/10.1007/JHEP06(2014)092).
- [78] ATLAS Collaboration. “In situ calibration of large-radius jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector”. In: *The European Physical Journal C* 79.2 (Feb. 2019), p. 005. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-6632-8](https://doi.org/10.1140/epjc/s10052-019-6632-8). URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6632-8>.
- [79] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. “A brief introduction to PYTHIA 8.1”. In: *Computer Physics Communications* 178.11 (June 2008), 852?867. ISSN: 0010-4655. DOI: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036). URL: <http://dx.doi.org/10.1016/j.cpc.2008.01.036>.
- [80] Stefano Carrazza, Stefano Forte, and Juan Rojo. *Parton Distributions and Event Generators*. 2013. arXiv: [1311.5887](https://arxiv.org/abs/1311.5887) [[hep-ph](#)].

- [81] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-021>. Geneva: CERN, Nov. 2014. URL: <https://cds.cern.ch/record/1966419>.
- [82] ATLAS Collaboration. *Performance of W/Z taggers using UFO jets in ATLAS*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-029>. Geneva: CERN, July 2021. URL: <http://cds.cern.ch/record/2777009>.
- [83] ATLAS Collaboration. “Jet mass reconstruction with the ATLAS Detector in early Run2 data”. In: (2016). ATLAS-CONF-2016-035.
- [84] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. “Energy Correlation Functions for Jet Substructure”. In: *JHEP* 06 (2013), p. 108. DOI: [10.1007/JHEP06\(2013\)108](https://doi.org/10.1007/JHEP06(2013)108). arXiv: [1305.0007 \[hep-ph\]](https://arxiv.org/abs/1305.0007).
- [85] Jesse Thaler and Ken Van Tilburg. “Identifying boosted objects with N-subjettiness”. In: *Journal of High Energy Physics* 2011.3 (Mar. 2011), pp. 134–136. ISSN: 1029-8479. DOI: [10.1007/jhep03\(2011\)015](https://doi.org/10.1007/jhep03(2011)015). URL: [http://dx.doi.org/10.1007/JHEP03\(2011\)015](http://dx.doi.org/10.1007/JHEP03(2011)015).
- [86] ATLAS Collaboration. “ATLAS measurements of the properties of jets for boosted particle searches”. In: *Physical Review D* 86.7 (Oct. 2012), 092?092. ISSN: 1550-2368. DOI: [10.1103/physrevd.86.072006](https://doi.org/10.1103/physrevd.86.072006). URL: <http://dx.doi.org/10.1103/PhysRevD.86.072006>.
- [87] Jesse Thaler and Lian-Tao Wang. “Strategies to identify boosted tops”. In: *Journal of High Energy Physics* 2008.07 (July 2008), 092?092. ISSN: 1029-8479. DOI: [10.1088/1126-6708/2008/07/092](https://doi.org/10.1088/1126-6708/2008/07/092). URL: <http://dx.doi.org/10.1088/1126-6708/2008/07/092>.
- [88] Chunhui Chen. “New approach to identifying boosted hadronically decaying particles using jet substructure in its center-of-mass frame”. In: *Physical Review D* 85.3 (Feb. 2012), 092?092. ISSN: 1550-2368. DOI: [10.1103/physrevd.85.034007](https://doi.org/10.1103/physrevd.85.034007). URL: <http://dx.doi.org/10.1103/PhysRevD.85.034007>.
- [89] Leandro G. Almeida et al. “Top quark jets at the LHC”. In: *Physical Review D* 79.7 (Apr. 2009), 092?092. ISSN: 1550-2368. DOI: [10.1103/physrevd.79.074012](https://doi.org/10.1103/physrevd.79.074012). URL: <http://dx.doi.org/10.1103/PhysRevD.79.074012>.
- [90] A. M. Sirunyan et al. “A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at 13 TeV”. In: *The European Physical Journal C* 80.3 (Mar. 2020), S08005. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-020-7773-5](https://doi.org/10.1140/epjc/s10052-020-7773-5). URL: <http://dx.doi.org/10.1140/epjc/s10052-020-7773-5>.
- [91] ATLAS Collaboration. “Search for WW,WZ resonance production in qq final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2018.3 (Mar. 2018), p. 8005. ISSN: 1029-8479. DOI: [10.1007/jhep03\(2018\)042](https://doi.org/10.1007/jhep03(2018)042). URL: [http://dx.doi.org/10.1007/JHEP03\(2018\)042](http://dx.doi.org/10.1007/JHEP03(2018)042).

- [92] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (July 2014), p. 8005. ISSN: 1029-8479. DOI: [10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079). URL: [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079).
- [93] Guido Altarelli, B. Mele, and M. Ruiz-Altaba. “Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders”. In: *Z. Phys. C* 45 (1989). [Erratum: *Z.Phys.C* 47, 676 (1990)], p. 109. DOI: [10.1007/BF01556677](https://doi.org/10.1007/BF01556677).
- [94] Charlotte Waltregny. *Improving diboson tagging efficiency for searches for new physics at the LHC*. URL: http://dpnc.unige.ch/MASTERS/MASTER_WALTREGNY.pdf.
- [95] Noemi Calace. “Jet Substructure Techniques for the Search of Diboson Resonances at the LHC and Performance Evaluation of the ATLAS Phase-II Inner Tracker Layouts”. Presented 20 Jun 2018. PhD thesis. Université de Genève, May 2018. URL: <http://cds.cern.ch/record/2628351>.
- [96] ATLAS Collaboration. *Selection of jets produced in 13TeV proton-proton collisions with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2015-029>. Geneva: CERN, July 2015. URL: <https://cds.cern.ch/record/2037702>.
- [97] J. M. Lindert et al. “Precise predictions for $V + \text{jets}$ dark matter backgrounds”. In: *The European Physical Journal C* 77.12 (Dec. 2017), p. 8005. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5389-1](https://doi.org/10.1140/epjc/s10052-017-5389-1). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5389-1>.
- [98] G. Choudalakis and D. Casadei. “Plotting the differences between data and expectation”. In: *The European Physical Journal Plus* 127.2 (Feb. 2012), p. 8005. ISSN: 2190-5444. DOI: [10.1140/epjp/i2012-12025-y](https://doi.org/10.1140/epjp/i2012-12025-y). URL: <http://dx.doi.org/10.1140/epjp/i2012-12025-y>.
- [99] 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: *Journal of High Energy Physics* 2013.9 (Sept. 2013), p. 300. ISSN: 1029-8479. DOI: [10.1007/jhep09\(2013\)076](https://doi.org/10.1007/jhep09(2013)076). URL: [http://dx.doi.org/10.1007/JHEP09\(2013\)076](http://dx.doi.org/10.1007/JHEP09(2013)076).
- [100] ATLAS Collaboration. *Search for diboson resonances with jets in 140 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*. Tech. rep. Geneva: CERN, Nov. 2018. URL: <https://cds.cern.ch/record/2647394>.
- [101] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ using the ATLAS detector at the LHC*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-021>. Geneva: CERN, June 2019. URL: <https://cds.cern.ch/record/2677054>.
- [102] Andriy Burkov. *The hundred-page machine learning book*. URL: <http://themlbook.com>. Andriy Burkov, 2019.
- [103] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. URL: <http://www.deeplearningbook.org>. MIT Press, 2016.

- [104] Ece Akilli. “Machine learning-based identification of boosted objects and search for pair production of heavy vector-like quarks in fully-hadronic final states with the ATLAS detector.” PhD thesis. Université de Genève, 2019. URL: <http://archive-ouverte.unige.ch/unige:138578>.
- [105] Harsh Pokharna. *Introduction to Neural Networks I*. URL: <https://medium.com/%5C%5C%20technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-c50f6012d5eb>.
- [106] Harsh Pokharna. *Introduction to Neural Networks II*. URL: <https://medium.com/%5C%5C%20technologymadeeasy/for-dummies-the-introduction-to-neural-networks-we-all-need-part-2-1218d5dc043>.
- [107] Harsh Pokharna. *Introduction to Convolutional Neural Networks*. URL: <https://medium.com/technologymadeeasy/the-best-explanation-of-convolutional-neural-networks-on-the-internet-fbb8b1ad5df8>.
- [108] ATLAS Collaboration. *Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-017>. Geneva: CERN, July 2017. URL: <https://cds.cern.ch/record/2275641>.
- [109] Yu-Chen Janice Chen et al. “Boosted W and Z tagging with jet charge and deep learning”. In: *Physical Review D* 101.5 (Mar. 2020), p. 455. ISSN: 2470-0029. DOI: [10.1103/physrevd.101.053001](https://doi.org/10.1103/physrevd.101.053001). URL: <http://dx.doi.org/10.1103/PhysRevD.101.053001>.
- [110] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: [1412.6980 \[cs.LG\]](https://arxiv.org/pdf/1412.6980.pdf). URL: <https://arxiv.org/pdf/1412.6980.pdf>.
- [111] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: [1502.03167 \[cs.LG\]](https://arxiv.org/pdf/1502.03167.pdf). URL: <https://arxiv.org/pdf/1502.03167.pdf>.
- [112] Xikun Zhang et al. *Graph Edge Convolutional Neural Networks for Skeleton Based Action Recognition*. 2018. arXiv: [1805.06184 \[cs.CV\]](https://arxiv.org/pdf/1805.06184.pdf). URL: <https://arxiv.org/pdf/1805.06184.pdf>.
- [113] Huilin Qu and Loukas Gouskos. “Jet tagging via particle clouds”. In: *Physical Review D* 101.5 (Mar. 2020), p. 455. ISSN: 2470-0029. DOI: [10.1103/physrevd.101.056019](https://doi.org/10.1103/physrevd.101.056019). URL: <http://dx.doi.org/10.1103/PhysRevD.101.056019>.
- [114] François Chollet et al. *Keras*. <https://keras.io>. 2015.
- [115] Micha Szleper. *The Higgs boson and the physics of WW scattering before and after Higgs discovery*. 2015. arXiv: [1412.8367 \[hep-ph\]](https://arxiv.org/pdf/1412.8367.pdf). URL: <https://arxiv.org/pdf/1412.8367.pdf>.
- [116] Carsten Bittrich. “Study of Polarization Fractions in the Scattering of Massive Gauge Bosons $W^\pm Z \rightarrow W^\pm Z$ with the ATLAS Detector at the Large Hadron Collider”. Presented 2015. PhD thesis. Technische Universität Dresden, Apr. 2015. URL: <https://cds.cern.ch/record/2014124>.

- [117] W. J. Stirling and E. Vryonidou. “Electroweak gauge boson polarisation at the LHC”. In: *Journal of High Energy Physics* 2012.7 (July 2012), p. 455. ISSN: 1029-8479. DOI: [10.1007/jhep07\(2012\)124](https://doi.org/10.1007/jhep07(2012)124). URL: [http://dx.doi.org/10.1007/JHEP07\(2012\)124](http://dx.doi.org/10.1007/JHEP07(2012)124).
- [118] Y. Gao. *Normalised differential distributions of the cross section over the cosine of the decay angle of an hypothetical new particle X*. Internal communication.
- [119] Ezio Maina. “Vector boson polarizations in the decay of the Standard Model Higgs”. In: *Physics Letters B* 818 (2021), p. 136360. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2021.136360>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269321003002>.
- [120] S. Agostinelli et al. “Geant4 a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [121] Xavier Glorot and Yoshua Bengio. *Understanding the difficulty of training deep feedforward neural networks*. 2010. URL: <https://proceedings.mlr.press/v9/glorot10a.html>.
- [122] Abien Fred Agarap. *Deep Learning using Rectified Linear Units (ReLU)*. 2019. arXiv: 1803.08375 [cs.NE]. URL: <https://arxiv.org/pdf/1803.08375.pdf>.
- [123] ATLAS Collaboration. “Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC”. In: *The European Physical Journal C* 79.5 (Apr. 2019), pp. 250–303. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-6847-8](https://doi.org/10.1140/epjc/s10052-019-6847-8). URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6847-8>.
- [124] S. Chatrchyan, A. Khachatryan, and et al. “Measurement of the Polarization of W Bosons with Large Transverse Momenta in $W + \text{jets}$ Events at the LHC”. In: *Phys. Rev. Lett.* 107 (2 July 2011), p. 021802. DOI: [10.1103/PhysRevLett.107.021802](https://doi.org/10.1103/PhysRevLett.107.021802). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.107.021802>.
- [125] ATLAS Collaboration. “Measurement of the W boson polarization in top quark decays with the ATLAS detector”. In: *Journal of High Energy Physics* 2012.6 (June 2012), p. 021802. ISSN: 1029-8479. DOI: [10.1007/jhep06\(2012\)088](https://doi.org/10.1007/jhep06(2012)088). URL: [http://dx.doi.org/10.1007/JHEP06\(2012\)088](http://dx.doi.org/10.1007/JHEP06(2012)088).