

# Graph Neural Network Flavour Tagging and Boosted Higgs Measurements at the LHC

Samuel John Van Stroud  
University College London

Submitted to University College London in fulfilment  
of the requirements for the award of the  
degree of **Doctor of Philosophy**

May 29, 2023

## **Declaration**

I, Samuel John Van Stroud confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Samuel Van Stroud

---

# Abstract

This thesis presents investigations into the challenges of, and potential improvements to,  $b$ -jet identification ( $b$ -tagging) at the ATLAS experiment at the Large Hadron Collider (LHC). The presence of  $b$ -jets is a key signature of many interesting physics processes such as the production of Higgs bosons, which preferentially decay to a pair of  $b$ -quarks. In this thesis, a particular focus is placed on the high transverse momentum regime, which is a critical region in which to study the Higgs boson and the wider Standard Model, but also a region within which  $b$ -tagging becomes increasingly difficult.

As  $b$ -tagging relies on the accurate reconstruction of charged particle trajectories (tracks), the tracking performance is investigated and potential improvements are assessed. Track reconstruction becomes increasingly difficult at high transverse momentum due to the increased multiplicity and collimation of tracks, and also due to the presence of displaced tracks from the decay of a long-flying  $b$ -hadron. The investigations reveal that the quality selections applied during track reconstruction are suboptimal for  $b$ -hadron decay tracks inside high transverse momentum  $b$ -jets, motivating future studies into the optimisation of these selections.

Two novel approaches are developed to improve  $b$ -tagging performance. Firstly, an algorithm which is able to classify the origin of tracks is used to select a more optimal set of tracks for input to the  $b$ -tagging algorithms. Secondly, a graph neural network (GNN) jet flavour tagging algorithm has been developed. This algorithm directly accepts jets and tracks as inputs, making a break from previous algorithms which relied on the outputs of intermediate

taggers. The model is trained to simultaneously predict the jet flavour, track origins, and the spatial track-pair compatibility, and demonstrates marked improvements in  $b$ -tagging performance both at low and high transverse momenta. The closely related task of  $c$ -jet identification also benefits from this approach.

Analysis of high transverse momentum  $H \rightarrow b\bar{b}$  decays, where the Higgs boson is produced in association with a vector boson, was performed using  $139\text{ fb}^{-1}$  of  $13\text{ TeV}$  proton-proton collision data from Run 2 of the LHC. This analysis provided first measurements of the  $VH$ ,  $H \rightarrow b\bar{b}$  process in two high transverse momentum regions, and is described with a particular focus on the background modelling studies performed by the author.

# Impact Statement

This thesis details research in experimental particle physics. The primary contributions are on the improvement of the data analysis algorithms which are used to process proton-proton collisions induced within the ATLAS detector at the Large Hadron Collider (LHC), and the analysis of candidate Higgs boson events.

The primary outcome of the research is an advancement of knowledge about how the Universe works on the most fundamental level, encoded for example in the improved measurement of the fundamental constants of the Standard Model, or in the observation of previously unseen particles or interactions. Although this kind of knowledge doesn't always have an immediate and direct relevance for society, potential applications are impossible to rule out and could have a very large impact further in the future, as has been seen with previous advancements in fundamental science.

The research does find indirect application in the form of associated technological developments that have transferable application within different fields. The cutting-edge techniques developed at CERN for ATLAS and the LHC have found many spin-off applications elsewhere in society, for example the World Wide Web, high-field magnet technology in MRI, touch-screen technology and cloud computing. Fundamental physics, as a method of proposing difficult and novel problems, can therefore be seen as a way to generate innovative technologies.

Working in the field also helps to train skilled researchers, which can be redeployed to other areas of society to tackle various problems. In this thesis advanced statistical and data science methods are

employed. Such methods currently find wide and varied use in many fields. The training of such highly skilled individuals has a sustained and significant positive economic impact.

Finally, the work carried at ATLAS and the LHC is widely publicised – support of and interest in fundamental physics research helps to generate excitement about science and technology, and educate people about how the Universe works. This in turn attracts people into the area, propagating the benefits described above.

## Acknowledgements

Firstly I give thanks to my supervisor Tim Scanlon for all the guidance and support he has offered over the course of this doctorate. Tim has always been consistent with clear explanations and sound advice throughout the last four years. I would also like to thank everyone I've worked with at ATLAS and at UCL. In particular I have Jonathan Shlomi to thank for the fruitful collaboration on advancements in flavour tagging. I would also like to thank Brian Moser and Hannah Arnold for their patient and kind support during the course of the  $VH$ ,  $H \rightarrow b\bar{b}$  analysis. I'm grateful to everyone I've worked with in the ATLAS Flavour Tagging and Tracking groups, in particular to Gabriel Facini, Sébastien Rettie, Valerio Dao, Bingxuan Liu and Francesco Di Bello for their guidance, and Dan Guest for his fastidious merge request reviews. This thesis was made in L<sup>A</sup>T<sub>E</sub>X 2 <sub>$\epsilon$</sub>  using the “heptesis” class [1].

# Contents

<b>1</b>	<b>Introduction</b>	<b>12</b>
<b>2</b>	<b>Theoretical Framework</b>	<b>14</b>
2.1	The Standard Model . . . . .	14
2.1.1	Quantum Electrodynamics . . . . .	15
2.1.2	Quantum Chromodynamics . . . . .	17
2.1.3	The Electroweak Sector . . . . .	18
2.2	The Higgs Mechanism . . . . .	19
2.2.1	Electroweak Symmetry Breaking . . . . .	20
2.2.2	Fermionic Yukawa Coupling . . . . .	23
2.2.3	Higgs Sector Phenomenology . . . . .	25
<b>3</b>	<b>The Large Hadron Collider and the ATLAS Detector</b>	<b>28</b>
3.1	The Large Hadron Collider . . . . .	29
3.2	Coordinate System & Collider Definitions . . . . .	30
3.2.1	ATLAS Coordinate System . . . . .	31
3.2.2	Track Parameterisation . . . . .	32
3.2.3	Hadron Collider Definitions . . . . .	32
3.3	The ATLAS Detector . . . . .	36
3.3.1	Inner Detector . . . . .	37
3.3.2	Calorimeters . . . . .	40
3.3.3	Muon Spectrometer . . . . .	43
3.3.4	The Trigger . . . . .	44
3.4	Reconstructed Physics Objects . . . . .	44
3.4.1	Tracks . . . . .	45
3.4.2	Vertices . . . . .	48
3.4.3	Jets . . . . .	49
3.4.4	Leptons . . . . .	52

3.4.5	Missing Transverse Momentum . . . . .	53
3.5	Flavour Tagging Algorithms . . . . .	54
3.5.1	Common Definitions . . . . .	54
3.5.2	Decay Topology . . . . .	55
3.5.3	Low-level Algorithms . . . . .	57
3.5.4	High-level Algorithms . . . . .	58
<b>4</b>	<b>Tracking and Flavour Tagging at High-<math>p_T</math></b>	<b>59</b>
4.1	Datasets . . . . .	59
4.2	$b$ -hadron Reconstruction Challenges . . . . .	61
4.3	High $p_T$ $b$ -hadron Tracking Improvements . . . . .	64
4.3.1	Shared Hits . . . . .	64
4.3.2	Global $\chi^2$ Fitter Outlier Removal . . . . .	65
4.4	Conclusion . . . . .	68
<b>5</b>	<b>Track Classification NN</b>	<b>70</b>
5.1	Machine Learning Background . . . . .	70
5.1.1	Neural Networks . . . . .	72
5.1.2	Training with Gradient Descent . . . . .	74
5.2	Track Truth Origin Labelling . . . . .	74
5.3	Fake Track Identification Tool . . . . .	75
5.3.1	Model Inputs . . . . .	77
5.3.2	Model Hyperparameters . . . . .	79
5.3.3	Results . . . . .	79
5.4	$b$ -hadron Track Identification . . . . .	81
5.5	Combined Approach . . . . .	81
5.6	Conclusion . . . . .	83
<b>6</b>	<b>Graph Neural Network Flavour Tagger</b>	<b>85</b>
6.1	Motivation . . . . .	85
6.2	Graph Neural Network Theory . . . . .	87
6.3	Experimental Setup . . . . .	89
6.3.1	Datasets . . . . .	89
6.4	Model Specification . . . . .	90
6.4.1	Model Inputs . . . . .	90
6.4.2	Auxiliary Training Objectives . . . . .	92

6.4.3	Architecture . . . . .	93
6.5	Training . . . . .	97
6.6	Results . . . . .	98
6.6.1	<i>b</i> -tagging Performance . . . . .	99
6.6.2	<i>c</i> -tagging Performance . . . . .	101
6.6.3	Ablations . . . . .	104
6.6.4	Inclusion of Low-Level Vertexing Algorithms . . . . .	106
6.6.5	Jet Display Diagrams . . . . .	106
6.6.6	Vertexing Performance . . . . .	108
6.6.7	Track Classification Performance . . . . .	113
6.7	Further Studies . . . . .	116
6.7.1	Looser Track Selection . . . . .	116
6.7.2	High Level Trigger . . . . .	117
6.7.3	High Luminosity LHC . . . . .	117
6.8	Conclusion . . . . .	120
<b>7</b>	<b>Boosted VHbb Analysis</b>	<b>124</b>
7.1	Analysis Overview . . . . .	125
7.1.1	Object Reconstruction . . . . .	126
7.1.2	Event Selection . . . . .	128
7.1.3	Control Regions . . . . .	129
7.1.4	Background Composition . . . . .	129
7.1.5	Data & Simulated Samples . . . . .	132
7.1.6	Overview of Systematic Uncertainties . . . . .	132
7.2	Experimental Uncertainties . . . . .	134
7.3	Modelling Uncertainties . . . . .	135
7.3.1	Truth Tagging . . . . .	136
7.3.2	Sources of Systematic Modelling Uncertainties . . . . .	136
7.3.3	Implementation of Variations . . . . .	138
7.3.4	Vector Boson + Jets Modelling . . . . .	139
7.3.5	Diboson Modelling . . . . .	145
7.3.6	$t\bar{t}$ and single-top Modelling . . . . .	147
7.3.7	Signal Modelling . . . . .	149
7.4	Statistical Treatment . . . . .	149
7.4.1	The Likelihood Function . . . . .	150

7.4.2	Background Normalisations	153
7.4.3	Asimov Dataset & Expected Results	154
7.5	Results	154
7.5.1	Post-fit Results	154
7.5.2	Observed Signal Strength & Significance	159
7.5.3	Impact of Systematics	161
7.5.4	STXS Interpretation	163
7.6	Conclusion	166
8	Conclusion	167
8.1	Summary	167
8.2	Future Work	168
<b>Bibliography</b>		170
<b>List of figures</b>		186
<b>List of tables</b>		198

# Chapter 1

## Introduction

This thesis describes efforts to improve the understanding of the Higgs boson and its coupling to heavy flavour quarks, primarily through the development of algorithms used to reconstruct and identify jets. The thesis is structured in the following manner:

Chapter 2 describes the theoretical foundations of the work presented in the rest of the thesis.

Chapter 3 describes the ATLAS detector and the CERN accelerator complex. Details of reconstructed physics objects and the  $b$ -jet identification algorithms are given.

Chapter 4 provides an overview of the challenges facing successful charged particle trajectory (track) reconstruction and correspondingly  $b$ -jet identification, with a particular focus on the high transverse momentum regime. Preliminary investigations into reconstruction improvements are provided.

Chapter 5 describes the development of an algorithm which predicts the origins of tracks. The tool is used to improve  $b$ -tagging performance by the identification and removal of fake tracks before their input to the  $b$ -tagging algorithms.

Chapter 6 introduces a novel, monolithic  $b$ -jet identification algorithm which makes use of graph neural networks and auxiliary training objectives.

Chapter 7 describes the measurement of the associated production of a Higgs boson decaying into a pair of  $b$ -quarks at high transverse momentum.

Chapter 8 contains some concluding remarks.

The author's contribution to the work presented in this thesis is as follows.

**Tracking:** The author was an active member of the Cluster and Tracking in Dense Environments group throughout their PhD, starting with their qualification task on the understanding of tracking performance at high transverse momentum (Chapter 4). The author played a key role in the validation for the tracking group of Release 22 of the ATLAS software, including the validation of the quasi-stable particle interaction simulation and the radiation damage Monte-Carlo simulation. The author helped design and improve several tracking software frameworks, and contributed to heavy flavour tracking efficiency studies in dense environments. The author developed a tool to identify and reject fake-tracks (Chapter 5), which is being investigated for use in the upcoming tracking paper.

***b*-tagging:** The author has been an active member of the Flavour Tagging group since October 2020. The author played a key role in investigating the performance of the low level taggers at high transverse momentum and led studies into the labelling and classification of track origins. Based on work by Jonathan Shlomi [2], the author helped develop a new flavour tagging algorithm which offers a large performance improvement with respect to the current state of the art (Chapter 6). The author was the primary editor of a public note associated with this work [3], which will also be further developed in an upcoming paper. The author also contributed to the proliferation of the new algorithm to the trigger, High Luminosity LHC, and  $X \rightarrow bb$  use cases. The author also played a key role in software r22 validation studies for the Flavour Tagging group, including the validation of the quasi-stable particle interaction simulation. The author maintains and contributes to various software frameworks used in the Flavour Tagging group, including as lead developer of three packages, to create training datasets, pre-process samples for performance studies and a framework for training graph neural networks, and contributes to group documentation.

**Higgs:** The author was an active member of the Boosted VHbb analysis group. The author performed various studies deriving systematic uncertainties for the  $V+jets$  and diboson backgrounds (Chapter 7). The author also produced and maintained samples, ran fit studies and cross checks, and gave the diboson unblinding approval talk to the Higgs group. The author also contributed to the development of the analysis software.

# Chapter 2

## Theoretical Framework

The Standard Model (SM) of particle physics [4, 5] is the theory describing all known elementary particles and their interactions via three of the four fundamental forces. Developed by merging the successful theories of quantum mechanics and special relativity in the second half of the 20th century, the SM's position today at the centre of our understanding of the nature of the Universe is firmly established by an unparalleled level of agreement between the model's predictions and experimental results [6, 7].

The SM has predicted the discovery of the top and bottom quarks [8–10], the  $W$  and  $Z$  bosons [11], and the tau neutrino [12]. The last missing piece of the SM to be discovered was the Higgs boson, first theorised in the 1960s [13–15], and eventually observed at the LHC in 2012 [16, 17]. After its discovery, much ongoing work has been carried out performing detailed measurements of its mass and interactions with other particles.

In this chapter, an overview of the SM is given in Section 2.1, and a more detailed discussion of the Higgs sector and Higgs phenomenology is provided in Section 2.2.

### 2.1 The Standard Model

The SM is formulated in the language of Quantum Field Theory (QFT). In this framework, particles are localised excitations of corresponding quantum fields, which are operator-valued distributions across spacetime.

Central to QFT is the Lagrangian density which describes the kinematics and dynamics of a field. Observations of conserved quantities are linked, via Noether's theorem, to symmetries which are expressed by the Lagrangian. Alongside Global Poincaré symmetry, the SM Lagrangian observes a local non-Abelian  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  gauge symmetry. Gauge symmetries leave observable properties of the system unchanged when the corresponding gauge transformations are applied to the fields. The full Lagrangian of the SM can be broken up into distinct terms corresponding to the different sectors, as in Eq. (2.1). An overview of each sector is given in the following chapters.

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} \quad (2.1)$$

The SM provides a mathematical description of how the electromagnetic, weak and strong forces interact with the matter content of the Universe. The particle content of the SM consists of spin-1/2 fermions, listed in Table 2.1, and integral spin bosons listed in Table 2.2.

<b>Generation</b>	Leptons			Quarks		
	<b>Flavour</b>	<b>Mass [MeV]</b>	<b>Charge [e]</b>	<b>Flavour</b>	<b>Mass [MeV]</b>	<b>Charge [e]</b>
First	$e$	0.511	-1	$u$	2.16	$\frac{2}{3}$
	$\nu_e$	$< 1.1 \times 10^{-6}$	0	$d$	4.67	$-\frac{1}{3}$
Second	$\mu$	105.7	-1	$c$	$1.27 \times 10^3$	$\frac{2}{3}$
	$\nu_\mu$	$< 0.19$	0	$s$	93.4	$-\frac{1}{3}$
Third	$\tau$	1776.9	-1	$t$	$173 \times 10^3$	$\frac{2}{3}$
	$\nu_\tau$	$< 18.2$	0	$b$	$4.18 \times 10^3$	$-\frac{1}{3}$

**Table 2.1:** The fermions of the SM [18]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.

### 2.1.1 Quantum Electrodynamics

Quantum electrodynamics (QED) is the relativistic quantum theory which describes the interactions between the photon and charged matter. Consider a Dirac spinor field  $\psi = \psi(x)$  and its adjoint  $\bar{\psi} = \psi^\dagger \gamma^0$ , where  $\psi^\dagger$  denotes the Hermitian conjugate

Name	Symbol	Mass [GeV]	Charge [ $e$ ]	Spin
Photon	$\gamma$	$< 1 \times 10^{-27}$	$< 1 \times 10^{-46}$	1
Charged Weak boson	$W^\pm$	$80.377 \pm 0.012$	$\pm 1$	1
Neutral Weak boson	$Z$	$91.1876 \pm 0.0021$	0	1
Gluon	$g$	0	0	1
Higgs	$H$	$125.25 \pm 0.17$	0	0

**Table 2.2:** The bosons of the SM [18]. The photon, weak bosons and gluons are gauge bosons arising from gauge symmetries, and carry the four fundamental forces of the SM. The recently discovered Higgs boson is the only fundamental scalar particle in the SM.

of  $\psi$ . The field  $\psi$  describes a fermionic spin-1/2 particle, for example an electron. The Dirac Lagrangian density is

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\cancel{\partial} - m)\psi, \quad (2.2)$$

where  $\cancel{\partial} = \gamma^\mu \partial_\mu$  denotes the contraction with the Dirac gamma matrices  $\gamma^\mu$  (summation over up-down pairs of indices is assumed). Application of the Euler-Lagrange equation on Eq. (2.2) yields the Dirac equation

$$(i\cancel{\partial} - m)\psi = 0. \quad (2.3)$$

Suppose some fundamental symmetry that requires invariance under a local  $U(1)$  gauge transformation

$$\psi \rightarrow \psi' = \psi e^{-iq\alpha(x)}, \quad (2.4)$$

where  $\alpha$  varies over every spacetime point  $x$ . Under this transformation, the Dirac equation transforms as

$$(i\cancel{\partial} - m)\psi e^{-iq\alpha(x)} + q\cancel{\partial}\alpha(x)\psi e^{-iq\alpha(x)} = 0. \quad (2.5)$$

For the Dirac equation to remain invariant under the transformation in Eq. (2.4), a new field  $A_\mu$  which transforms as  $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x)$  must be added. The transformed interaction term

$$-q\cancel{A}\psi \rightarrow -q\cancel{A}\psi e^{-iq\alpha(x)} - q\cancel{\partial}\alpha(x)\psi e^{-iq\alpha(x)} \quad (2.6)$$

will then cancel the asymmetric term in Eq. (2.5) as required. The  $U(1)$  invariant Lagrangian can therefore be constructed by adding an interaction between the  $\psi$  and  $A_\mu$  fields to Eq. (2.2). For completeness, the kinetic term for the new field  $A_\mu$  is also added in terms of  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ , which is trivially invariant under the transformation in Eq. (2.4). The interaction term is typically absorbed into the covariant derivative  $D_\mu = \partial_\mu + iqA_\mu$ , thus named as it transforms in the same way as the field  $\psi$ . Collecting these modifications to Eq. (2.2) yields the QED Lagrangian

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(iD^\mu - m)\psi. \quad (2.7)$$

Requiring invariance under local  $U(1)$  gauge transformations necessitated the addition of a new field  $A_\mu$ , interpreted as the photon field, which interacts with charged matter. The quadratic mass term  $A_\mu A^\mu$  is not invariant and therefore the field  $A_\mu$  is required to be massless. In the SM, the QED Lagrangian is absorbed into the electroweak sector, discussed in Section 2.1.3.

### 2.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the study of quarks, gluons and their interactions. Quarks and gluons carry colour charge, which comes in three kinds, called red, green and blue. While the  $U(1)$  symmetry group in Section 2.1.1 was Abelian, the QCD Lagrangian is specified by requiring invariance under transformations from the non-Abelian  $SU(3)$  group, making it a Yang-Mills theory [19] which requires the addition of self-interacting gauge fields. The infinitesimal  $SU(3)$  group generators are given by  $T_a = \lambda_a/2$ , where  $\lambda_a$  are the eight Gell-Mann matrices. These span the space of infinitesimal group transformations and do not commute with each other, instead satisfying the commutation relation

$$[T_a, T_b] = if_{abc}T_c, \quad (2.8)$$

where  $f_{abc}$  are the group's structure constants. Consider the six quark fields  $q_k = q_k(x)$ . Each flavour of quark  $q_k$  transforms in the fundamental triplet representation, in which each component of the triplet corresponds to the colour quantum number for red, green and blue colour charged respectively.  $G_{\mu\nu}^a$  are the eight gluon field

strength tensors, one for each generator  $T_a$ , defined as

$$G_{\mu\nu}^a = \partial_\mu A_\nu - \partial_\nu A_\mu - g_s f^{abc} A_\mu^b A_\nu^c, \quad (2.9)$$

where  $A_\mu^a$  are the gluon fields and  $g_s$  is the strong coupling constant. The covariant derivative is written as

$$D_\mu = \partial_\mu + ig_s T_a A_\mu^a. \quad (2.10)$$

The full QCD Lagrangian is then given by

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} + q_k (i \not{D} - m_k) q_k. \quad (2.11)$$

Cubic and quartic terms of the gauge fields  $A_\mu^a$  appear in the Lagrangian, leading to the gluon's self interaction.

The QCD coupling constant  $g_s$  varies, or “runs”, with energy. At lower energy scales (and corresponding larger distance scales) the interaction is strong. This leads to quark confinement, whereby an attempt to isolate individual colour-charged quarks requires so much energy that additional quark-antiquark pairs are produced. At higher energy scales (and corresponding smaller distance scales), asymptotic freedom occurs as the interactions become weaker, allowing perturbative calculations to be performed. Hadrons are bound states of quarks. They are invariant under  $SU(3)$  gauge transformations (i.e. are colour-charge neutral, or *colourless*).

### 2.1.3 The Electroweak Sector

The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam (GWS) model of electroweak interaction [20–22]. The Lagrangian is specified by requiring invariance under the symmetry group  $SU(2)_L \otimes U(1)_Y$ , as motivated by a large amount of experimental data [23–25]. Here,  $SU(2)_L$  is referred to as weak isospin and  $U(1)_Y$  as weak hypercharge.

The generators of  $SU(2)_L$  are  $T_a = \sigma_a/2$ , where  $\sigma_a$  are the three Pauli spin matrices which satisfy the commutation relation

$$[T_a, T_b] = i\varepsilon_{abc}T_c. \quad (2.12)$$

The generator of  $U(1)_Y$  is  $Y = 1/2$ . Each generator corresponds to a gauge field, which, after symmetry breaking (discussed in Section 2.2), give rise to the massive vector bosons,  $W^\pm$  and  $Z$ , and the massless photon. The massive vector bosons are the carriers of the weak force. Due to the mass of the force carriers, the weak force has a short range and so it appears weak even though its intrinsic strength is comparable to that of QED.

The charge operator  $Q$  can be written as a combination of the third  $SU(2)_L$  generator and the  $U(1)_Y$  generator as in

$$Q = T_3 + Y. \quad (2.13)$$

The weak force violates parity conservation, i.e. weak interactions are not invariant under parity transformations (mirror reflections). Only left handed fermions participate in the weak interaction. Since there is no other force through which neutrinos interact with other particles, there are no right handed neutrinos in the Standard Model.

## 2.2 The Higgs Mechanism

The Brout-Englert-Higgs mechanism (henceforth just the “Higgs mechanism”) is the process through which the fundamental particles of the SM acquire mass [13–15]. Experimentally it was known that the weak force had a low effective strength, which was suggestive of a massive mediating gauge particle. However, directly adding mass to the weak gauge bosons violates the non-Abelian symmetry of the SM. Instead, the gauge bosons acquire mass through their interaction with a scalar Higgs doublet during the spontaneous breakdown of symmetry as discussed in Section 2.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as discussed in Section 2.2.2. Section 2.2.3 described some basic phenomenology of the Higgs particle relevant to hadron colliders.

### 2.2.1 Electroweak Symmetry Breaking

Spontaneous symmetry breaking (SSB) is a key part of the Higgs mechanism. It is the transition of a physical system from a state of manifest symmetry to a state of hidden, or *broken*, symmetry. In particular, this applies to physical systems where the Lagrangian observes some symmetry, but the lowest energy vacuum states do not exhibit that same symmetry. In other words, the symmetry is broken for perturbations around the vacuum state.

Consider the case in which the gauge fields from the local  $SU(2)_L \otimes U(1)_Y$  symmetry group (discussed in Section 2.1.3) are coupled to a complex scalar field  $\phi = \phi(x)$ , which transforms as a weak isospin doublet. Omitting the kinetic term of the gauge fields, and writing  $\phi^2 \equiv \phi^\dagger \phi$ , the Lagrangian is

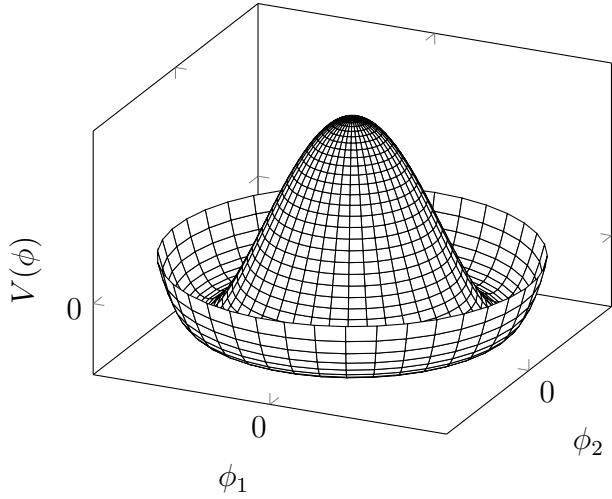
$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - \left[ \mu^2 \phi^2 + \frac{1}{2} \lambda^2 \phi^4 \right], \quad (2.14)$$

where the covariant derivative is given by

$$D_\mu = \partial_\mu + igA_\mu^a T^a + ig'B_\mu, \quad (2.15)$$

and  $T^a$  are the generators of  $SU(2)$ . The potential term  $V(\phi)$  is made up of a quadratic and quartic term in the scalar field  $\phi$ , which each contain an arbitrary parameter, respectively  $\lambda$  and  $\mu$ . The quartic term gives the field self-interaction, and cannot be negative as this would lead to a potential that was unbounded from below. The quadratic term can be positive or negative. In the case where the quadratic term is positive, it is interpreted as a mass term for the scalar field. By choosing  $\mu^2 < 0$  the field becomes unphysical due to its negative mass. The shape of the potential in this case is shown in Fig. 2.1. Note that in the case of the Standard Model, the scalar field  $\phi$  is a complex doublet, and so the corresponding potential is 5-dimensional.

In order to obtain a physical interpretation of the Lagrangian in Eq. (2.14) for the case where  $\mu^2 < 0$ , the field  $\phi$  is expanded around the vacuum state. The vacuum expectation value (VEV) is the expected value of the field  $\phi$  which minimises the potential  $V(\phi)$  (equivalently the expected value of the field operator  $\phi$  when the system is in a vacuum state,  $|\langle \phi \rangle_0|^2 \equiv |\langle 0 | \phi | 0 \rangle|^2 \equiv \phi_0^2$ ). Minimising the potential



**Figure 2.1:** The Higgs potential  $V(\phi)$  of the complex scalar field singlet  $\phi = \phi_1 + i\phi_2$ , with a choice of  $\mu^2 < 0$  leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin.

yields a VEV of

$$\phi_0^2 = -\mu^2/\lambda = v^2. \quad (2.16)$$

Due to the shape of the potential in Fig. 2.1, there is a degeneracy in the direction that the complex doublet  $\phi$  points. As all the different vacuum states minimise the potential and therefore yield identical physics, one can arbitrarily choose the state to lie along the second component of the doublet. Application of Eq. (2.13) shows this choice is manifestly invariant under the charge operator. This allows the identification of the unbroken subgroup  $U(1)_Q$ , under which the ground state is invariant. The generator of  $U(1)_Q$  is the charge operator  $Q$ .

Adding the particle content back to the theory by expanding the field around the vacuum state, and making a transformation to the unitary gauge to remove unphysical Nambu-Goldstone modes (which arise in the context of global symmetries [26, 27]), yields

$$\phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v + H \end{bmatrix}, \quad (2.17)$$

where  $H$  is a real scalar field, the *true vacuum* Higgs field. Substituting this into Eq. (2.14) and identifying physical fields from the quadratic terms of linear combinations of unphysical fields, one can write the physical fields  $W_\mu^\pm$ ,  $Z_\mu$  and  $A_\mu$  in terms of the original fields  $A_\mu^a$  and  $B_\mu$ . This gives

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2) \quad \begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix} = \begin{bmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{bmatrix} \begin{bmatrix} B_\mu \\ A_\mu^3 \end{bmatrix}. \quad (2.18)$$

where  $\theta_W$  is the weak mixing angle defined by

$$\cos \theta_W = \frac{g}{\sqrt{g^2 + g'^2}}. \quad (2.19)$$

The corresponding masses of the now massive vector bosons can be read off as

$$m_W = \frac{1}{2}gv \quad m_Z = \frac{m_W}{\cos \theta_W}, \quad (2.20)$$

while the photon remains massless. The Higgs mass is  $m_H = v\sqrt{\lambda} = \mu$ .

This is the Higgs mechanism. It maintains the renormalisability and unitarity of the SM whilst allowing the weak vector bosons to acquire mass. In summary, an unphysical complex scalar field  $\phi$  with a nonzero VEV leads to spontaneous symmetry breaking. Due to the non-Abelian symmetry breaking, would-be massless Nambu-Goldstone modes, which arise after expansion around the true vacuum state, are cancelled out by making a local gauge transformation to the unitary gauge, and instead are absorbed by the vector bosons, allowing them to acquire mass.

This sector of the SM contains four fundamental parameters that must be determined from experiment. These can be specified by the Lagrangian parameters  $g$ ,  $g'$ ,  $v$  and  $\lambda$  or the physically measurable parameters  $m_Z$ ,  $\sin \theta_W$ ,  $m_H$  and  $e$ . In the local neighbourhood around the true vacuum, the macroscopic symmetry of the system is not realised, and therefore the physical particles do not obey the original symmetry. However, information about the symmetry is retained through some additional constraints on the parameters of the theory. Prior to symmetry breaking, the potential contained two terms and two constants. After symmetry breaking there are three terms but still only two constants that relate these terms. This is the vestige of the original symmetry.

Spontaneous symmetry breaking has modified the original symmetry group of the SM  $SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \otimes U(1)_Q$ . Three broken generators from the symmetry group  $SU(2)_L \times U(1)_Y$  have been absorbed into the definition of the physical weak vector bosons, giving them mass. The same methodology can be used to generate the fermion masses, as shown in the next section.

### 2.2.2 Fermionic Yukawa Coupling

Adding the masses of the fermions by hand breaks the gauge invariance of the theory. Instead, we can use a Yukawa coupling between the fermion fields and the Higgs field in order to generate mass terms after spontaneous electroweak symmetry breakdown [21]. In this way, the fermion masses are determined by both the respective couplings to the Higgs field and the VEV of the Higgs field itself, which sets the basic mass scale of the theory.

The Higgs field  $\phi$  transforms as an  $SU(2)$  doublet with  $Y = 1/2$ , as does the left-handed fermion field  $\psi_L$ . The right-handed fermion field  $\psi_R$  transforms as an  $SU(2)$  singlet.

#### Charged Lepton Masses

The renormalisable and gauge invariant coupling between a fermionic field  $\psi$  and a scalar Higgs field  $\phi$  can be written as

$$\mathcal{L}_{\text{Yukawa}} = -g_f(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L). \quad (2.21)$$

where  $\psi_L = (\nu_L, e_L)$  and  $\psi_R = e_R$  for the first generation leptons. After spontaneous symmetry breaking (see Section 2.2.1), the scalar Higgs field in unitary gauge Eq. (2.17) consists of a VEV and the true vacuum Higgs field  $H$ . Substituting this in to Eq. (2.21) yields

$$\mathcal{L}_{\text{Yukawa}} = -\frac{v g_e}{\sqrt{2}} \bar{e} e - \frac{g_e}{\sqrt{2}} \bar{e} H e, \quad (2.22)$$

using  $\bar{e} e = (\bar{e}_L + \bar{e}_R)(e_L + e_R) = \bar{e}_L e_R + \bar{e}_R e_L$ . The VEV component of  $\phi$  provides the first term in Eq. (2.22) which is quadratic in the electron field, and can therefore

be identified as the electron mass term. An interaction term between the electron field  $e$  and the true vacuum Higgs field  $H$  is also present. Mass is generated for the other charged lepton generations in the same way.

### Quark Masses

The down-type quarks acquire their mass analogous to the leptons, with  $\psi_L = (u_L, d_L)$  and  $\psi_R = d_R$  for the first quark generation. Mass is generated for the up-type quarks using the conjugate field to  $\phi$  which transforms under  $SU(2)$  as a doublet with  $Y = -1/2$ . The conjugate field  $\tilde{\phi}$  is constructed as

$$\tilde{\phi} = i\sigma_2 \phi^\dagger = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1^\dagger \\ \phi_2^\dagger \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} v + H \\ 0 \end{bmatrix}, \quad (2.23)$$

and transforms in the same way as  $\phi$ . This field can be used to write an additional Yukawa coupling which provides mass for the up-type quarks in a similar way as before.

$$\mathcal{L}_{\text{Yukawa}}^{\text{up}} = -g_q (\bar{\psi}_L \tilde{\phi} \psi_R + \bar{\psi}_R \tilde{\phi}^\dagger \psi_L) \quad (2.24)$$

Considering the first generation of up-type quarks with  $\psi_L = (u_L, d_L)$  and  $\psi_R = u_R$ , substitution into Eq. (2.24) yields

$$\mathcal{L}_{\text{Yukawa}}^{\text{up}} = -\frac{vg_u}{\sqrt{2}} \bar{u}u - \frac{g_u}{\sqrt{2}} \bar{u}Hu. \quad (2.25)$$

The Yukawa terms mix quarks of different generations. Physical particles are detected in their mass eigenstates  $q$ , which diagonalise the mass matrix, but interact via the weak interaction according to their weak eigenstates  $\tilde{q}$ , which are superpositions of the mass eigenstates. This feature of the weak sector leads to mixing between different generations of quarks. Quark mixing can be expressed using the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which specifies the strength of flavour-changing weak currents. The entries in the matrix are enumerated as

$$\begin{bmatrix} \tilde{d} \\ \tilde{s} \\ \tilde{b} \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}, \quad (2.26)$$

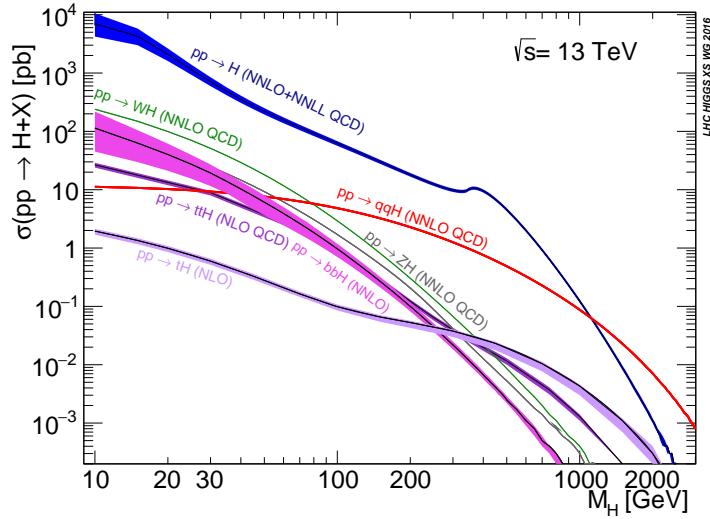
where the size of the elements  $|V_{pq}|^2$  measures the probability of a transition between states  $p$  and  $q$ .

### 2.2.3 Higgs Sector Phenomenology

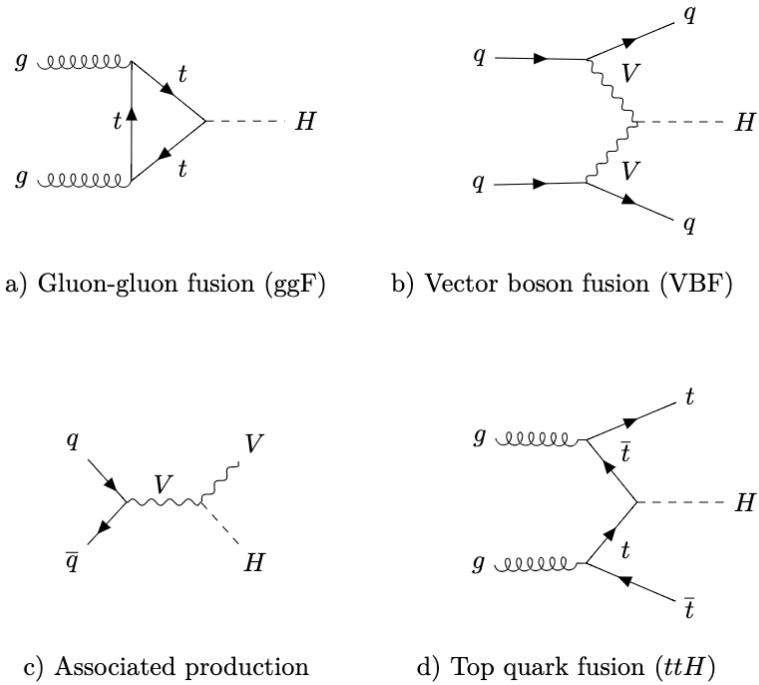
As previously discussed in this chapter, the Higgs field plays a key role in the SM by giving mass to fundamental particles. The strength of the coupling between the Higgs field and another particle is proportional to that particle's mass. This fact dictates which production mechanisms and decay modes are dominant at the LHC. The cross sections for different production mechanisms at a centre of mass energy  $\sqrt{s} = 13\text{ TeV}$  are shown as a function of the Higgs mass  $m_H$  in Fig. 2.2. At leading order in QCD, Higgs boson production occurs mainly through four modes, shown in Fig. 2.3. The dominant production mode is gluon-gluon fusion ( $pp \rightarrow H$ ), which is predominantly mediated by a virtual top quark loop. Vector boson fusion ( $pp \rightarrow qqH$ ) is the second most likely production mechanism, in which a pair of  $W$  or  $Z$  bosons fuse to produce a Higgs after being radiated by two quarks. Next most common is the associated production of a Higgs boson and a vector boson ( $pp \rightarrow VH$ ), in which a pair of quarks fuse to produce a single  $W$  or  $Z$  boson which radiates a Higgs. The final of the four leading production modes is top quark fusion ( $pp \rightarrow ttH$ ), in which two gluons each radiate a quark-antiquark pair, and a quark from each pair fuses to produce a Higgs boson.

Although gluon-gluon fusion is the dominant production mode, for hadronic decays of the Higgs boson the associated production with a vector boson has the advantage of leading to a more distinct final state due to the likelihood of the vector bosons decaying to leptons. Leptons provide clean signals to detect and trigger on.

Since the Higgs boson couples proportional to mass, decays to heavier particles are favoured. The branching ratios of different Higgs boson decay modes are shown as a function of  $m_H$  in Fig. 2.4. Approximately 58% of the time the Higgs boson decays to a pair of  $b$ -quarks, the dominant decay mode. The next most likely decay mode is to a pair of  $W$  bosons, which occurs approximately 20% of the time. After the  $b$ -quark, the next heaviest fermions are the tau lepton and the  $c$ -quark, decays to pairs of these particles happen approximately an order of magnitude less often. Decays to pairs of vector bosons are via a virtual off shell Higgs boson only. While the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ$  branching ratios are small compared with fermionic decay

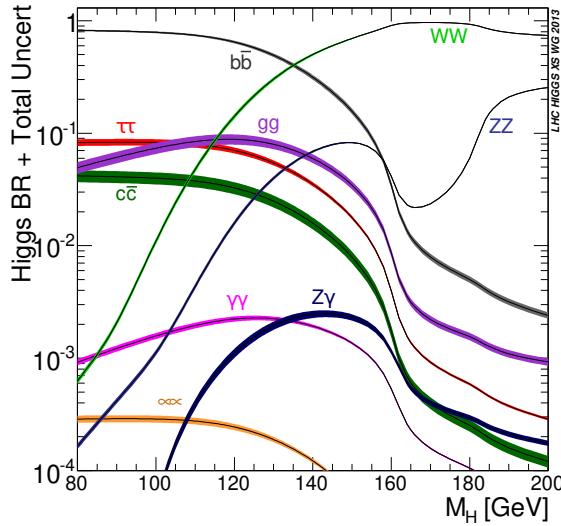


**Figure 2.2:** Higgs boson production cross sections as a function of Higgs mass ( $m_H$ ) at  $\sqrt{s} = 13 \text{ TeV}$  [28]. Uncertainties are shown in the shaded bands. At  $m_H = 125 \text{ GeV}$ , Higgs boson production is dominated by gluon-gluon fusion, vector boson fusion, associated production with vector bosons, and top quark fusion.



**Figure 2.3:** Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass  $m_H = 125 \text{ GeV}$  at a centre of mass energy  $\sqrt{s} = 13 \text{ TeV}$ .

modes (around 0.2% for  $H \rightarrow \gamma\gamma$ ), these decay channels were instrumental in the initial discovery of the Higgs due to the low level of background processes which mimic the final state [16, 17].



**Figure 2.4:** Higgs boson branching ratios as a function of Higgs mass ( $m_H$ ) at  $\sqrt{s} = 13$  TeV [28]. Uncertainties are shown in the shaded bands. At  $m_H = 125$  GeV, the Higgs predominantly decays to a pair of  $b$ -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of  $W$  bosons.

This thesis presents a measurement of the cross section of Higgs bosons produced in association with a vector boson and decaying to a pair of  $b$ -quarks, i.e.  $pp \rightarrow VH, H \rightarrow b\bar{b}'$ . The  $H \rightarrow b\bar{b}'$  decay mode directly probes the Higgs coupling to fermions, and more specifically to the bottom quark. This coupling was first observed in 2018 [29, 30]. Ongoing work measuring the coupling strengths, in particular in the high energy regime, is the focus of the analysis presented in this thesis in Chapter 7.

# Chapter 3

## The Large Hadron Collider and the ATLAS Detector

The Large Hadron Collider (LHC) at CERN is the world’s largest and most powerful particle accelerator [31]. Since the completion of its construction in 2008, the LHC has extended the frontiers of particle physics through significant increases in centre of mass energy and collision rates as compared with previous collider experiments. The LHC accelerates bunches of protons around a 27 km ring until they are travelling just  $3\text{ m s}^{-1}$  slower than the speed of light, at which point they are made to collide. The proton bunches travel round the ring 11,000 times per second in two concentric beams, which are guided by superconducting magnets cooled using liquid helium to  $-271.3\text{ }^{\circ}\text{C}$  ( $1.9\text{ K}$ ). The beams travel in opposite directions around the ring and are crossed at four locations so that collisions between protons can take place. Around these collision points four specialised detectors, ALICE [32], CMS [33], LHCb [34] and ATLAS [35], are located to capture information about the products of the collisions.

The work in this thesis is based on simulated and observed proton-proton collision data from the ATLAS detector. In this chapter, a brief overview of the LHC and the accelerator complex at CERN is given in Section 3.1. The coordinate system used at the ATLAS detector and other common definitions are introduced in Section 3.2. In Section 3.3 the different ATLAS detector systems are described, and in Section 3.4 information about various commonly used reconstructed objects is given. Finally, Section 3.5 provides an introduction to flavour tagging at ATLAS.

### 3.1 The Large Hadron Collider

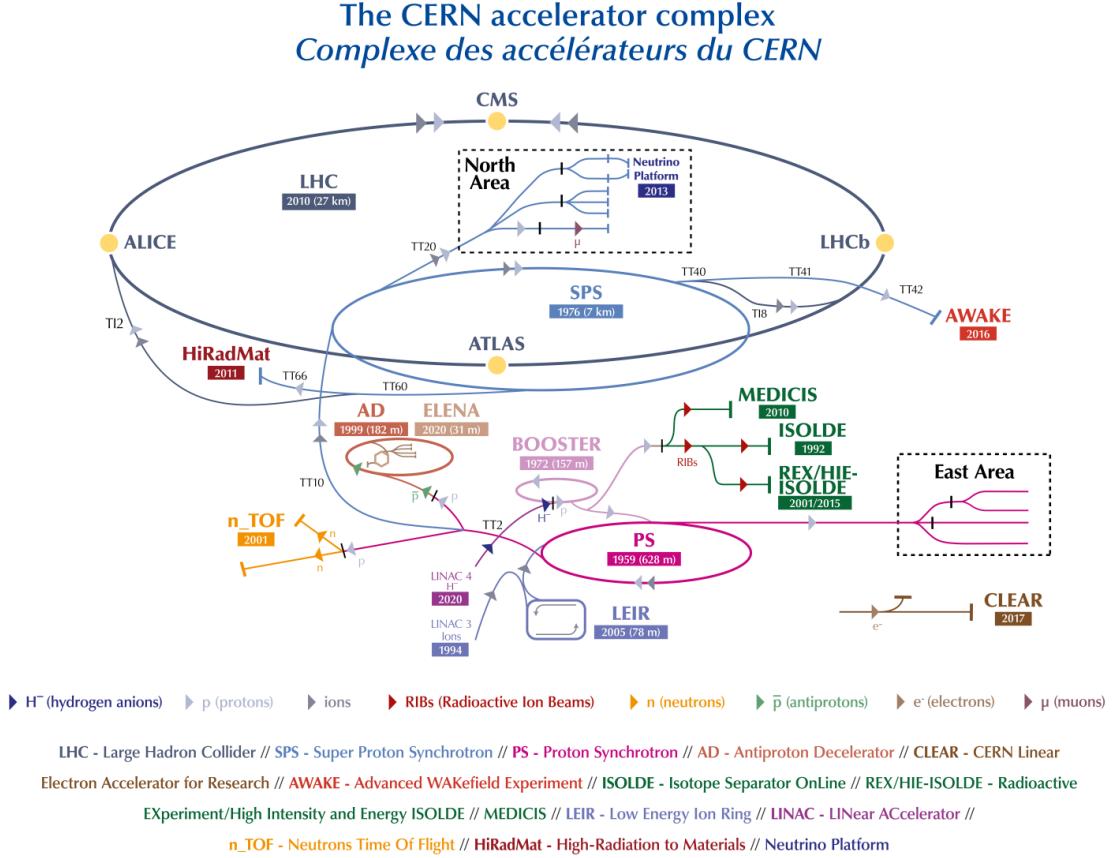
The LHC is operated in multi-year *runs* during which beams of protons are circulated and collided. Between runs there are periods of shutdown while the accelerator and detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC collided proton bunches, each containing more than  $10^{11}$  particles, 20 million times per second, providing 7 TeV proton-proton collisions at instantaneous luminosities of up to  $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$  (see Section 3.2.3). The centre-of-mass energy was increased to 8 TeV in 2012. Over the course of Run 1,  $26.4 \text{ fb}^{-1}$  of usable integrated luminosity was recorded. Run 2, which spanned 2015–2018, further increased the proton-proton collision energy to 13 TeV. During Run 2 the bunch spacing was reduced, leading to a collision rate of 40 MHz. Over the course of Run 2 a total usable integrated luminosity of  $139 \text{ fb}^{-1}$  was recorded. 2022 marked the beginning of Run 3 which, with a higher center of mass energy and peak instantaneous luminosity, is expected to culminate in an approximate tripling of the dataset size. A summary of key information about each run is listed in Table 3.1.

Period	Year	$\sqrt{s}$ [TeV]	$\langle\mu\rangle$	Bunch spacing [ns]	Luminosity [ $\text{cm}^{-2} \text{ s}^{-1}$ ]
Run 1	2010–2012	7–8	18	50–150	$8 \times 10^{33}$
Run 2	2015–2018	13	34	25	$1\text{--}2 \times 10^{34}$
Run 3	2022–2025	13.6	50	25	$2 \times 10^{34}$

**Table 3.1:** Overview of the different LHC runs [36,37]. The average number of interactions per bunch-crossing is denoted as  $\langle\mu\rangle$  (see Section 3.2.3), and is here averaged over the entire run. The luminosity is the peak instantaneous luminosity. Numbers for Run 3 are preliminary and are only provided to give an indication of expected conditions.

An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is at the final stage of a chain of accelerators which incrementally step-up the energy of incoming protons. The first accelerator is Linac2 (which has been replaced by Linac4 in 2020), a linear accelerator which accelerates negative hydrogen ions to an energy of 160 MeV. Upon leaving Linac4, the ions are stripped of both electrons and the bare protons are fed into the Proton Synchrotron Booster (PSB), which increases the energy of the protons to 2 GeV. The protons leaving the PSB are passed to the Proton Synchrotron (PS), which increases the energy to 26 GeV, and then from the PS to the Super Proton Synchrotron (SPS) which further increases the

energy to 450 GeV. Finally, the proton beams are injected in the LHC where they are accelerated to their final energy of 6.5 TeV (for Run 2).



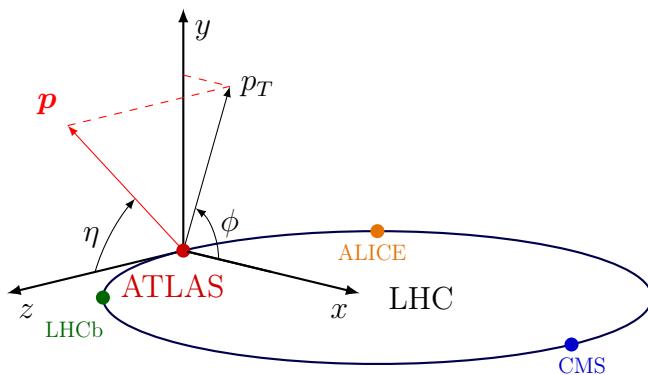
**Figure 3.1:** An overview of the CERN accelerator complex [38]. The LHC is fed by a series of accelerators starting with Linac2 (or Linac4 from 2020). Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

## 3.2 Coordinate System & Collider Definitions

In Section 3.2.1, the coordinate system used at ATLAS is introduced. The parameterisation used for specifying the trajectory of charged particle tracks is described in Section 3.2.2, and definitions for some frequently occurring concepts and quantities is provided in Section 3.2.3.

### 3.2.1 ATLAS Coordinate System

The origin of the coordinate system used by ATLAS is the nominal interaction point in the centre of the detector. As shown in Fig. 3.2, the  $z$ -axis points along the direction of the beam pipe, while the  $x$ -axis points from the interaction point to the centre of the LHC ring, and the  $y$ -axis points upwards. The transverse plane lies in  $x$ - $y$  while the longitudinal plane lies along the  $z$ -axis. A cylindrical coordinate system with coordinates  $(r, \phi)$  is used in the transverse plane, where  $r$  is the radius from the origin and  $\phi$  is the azimuthal angle around the  $z$ -axis.



**Figure 3.2:** The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum  $\mathbf{p} = (p_x, p_y, p_z)$  is shown by the red arrow. Reproduced from Ref. [39].

The polar angle  $\theta$  is commonly specified in terms of the pseudorapidity  $\eta$ , defined as

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]. \quad (3.1)$$

The transverse momentum  $p_T$  of an object is the sum in quadrature of the momenta in the transverse plane

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (3.2)$$

Angular distance between two objects is measured in units of  $\Delta R$  and is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \quad (3.3)$$

where  $\Delta\eta$  and  $\Delta\phi$  are the differences respectively in pseudorapidity and azimuthal angle between the two objects.

### 3.2.2 Track Parameterisation

The reconstructed trajectory of a charged particle (called a *track*) is parameterised as a helix which is fully specified using five parameters:  $(d_0, z_0, \phi, \theta, q/p)$ . The transverse and longitudinal impact parameters (IP)  $d_0$  and  $z_0$  specify the closest approach of the trajectory of a particle to a given origin, where the hard scatter primary vertex (see Section 3.4.2) is used in this thesis. The parameters  $\phi$  and  $\theta$  are the azimuthal and polar angles respectively, and  $q/p$  is the measured charge on the track<sup>1</sup> divided by the scalar 3-momentum. Fig. 3.3 shows each of these parameters diagrammatically.

Impact parameter significances are defined as the IP divided by its corresponding uncertainty  $\sigma$ , as in  $s(d_0) = d_0/\sigma(d_0)$  and  $s(z_0) = z_0/\sigma(z_0)$ . When used in flavour tagging (see Chapter 4), track IP significances are lifetime signed according to the track’s direction with respect to the jet axis and the primary vertex [41]. The signed IP significances are positive if the track crosses the jet axis in front of the primary vertex and negative if the track crosses behind the primary vertex.

### 3.2.3 Hadron Collider Definitions

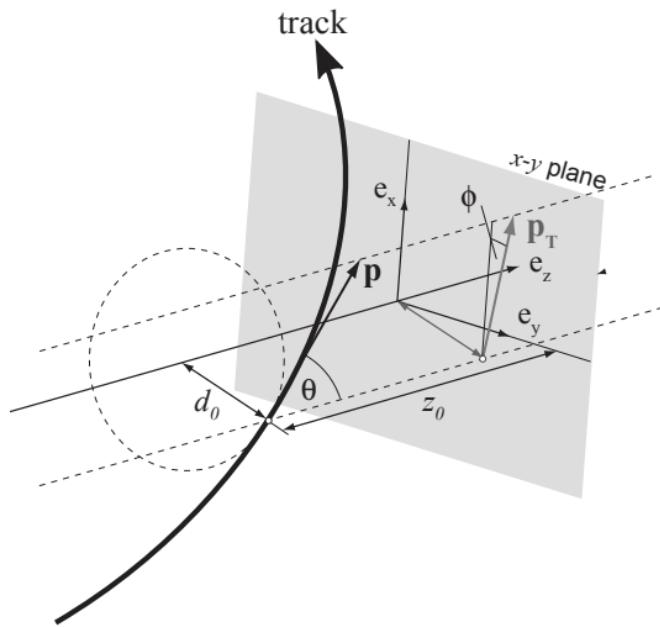
#### Cross Section

The cross section  $\sigma$  is a measure of the probability of an interaction between two colliding particles, and is analogous to an effective cross-sectional area of the particles. The cross section of a process depends on the transition matrix element and a phase space integral. At hadron colliders such as the LHC, the proton-proton cross section can be factorised as

$$\sigma(pp \rightarrow X) \approx \text{PDFs} \cdot \text{partonic cross section}. \quad (3.4)$$

---

<sup>1</sup>Reconstructed charged particles are assumed to have a charge of  $\pm 1$ .



**Figure 3.3:** The track parameterisation used at the ATLAS detector. Five coordinates ( $d_0, z_0, \phi, \theta, q/p$ ) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum  $\mathbf{p}$  and the transverse momentum  $p_T$  (defined in Eq. (3.2)). The basis vectors  $e_x$ ,  $e_y$  and  $e_z$  are also shown. Reproduced from Ref. [40].

The partonic cross section can be calculated for the high energy interactions at the LHC, while the parton distribution functions (PDFs), used to describe lower energy phenomena, have to be extracted from experimental results.

## Luminosity

The total number of proton-proton collisions  $N$  is related to the total proton-proton cross section  $\sigma$  by the integrated luminosity  $L$ , as in

$$N = \sigma L = \sigma \int \mathcal{L} dt. \quad (3.5)$$

The instantaneous luminosity  $\mathcal{L}$  relates the cross section to the number of collisions per unit time. For two colliding bunched proton beams, it is defined as

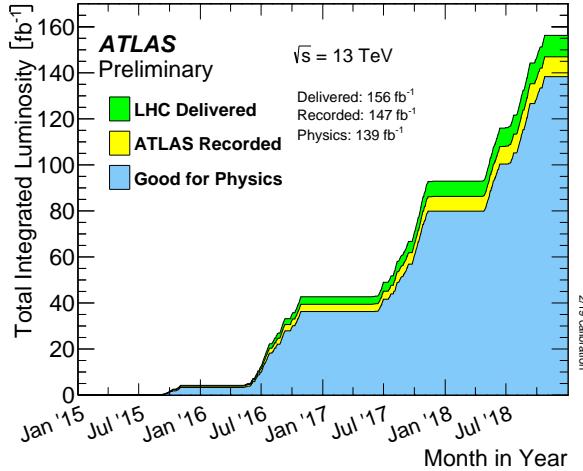
$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} = \frac{fn_1n_2}{4\pi\sigma_x\sigma_y}, \quad (3.6)$$

where  $n_1$  and  $n_2$  are the number of protons in the colliding bunches,  $f$  is the bunch crossing frequency, and  $\sigma_x$  and  $\sigma_y$  are the RMS width of the beam in the horizontal and vertical directions respectively.

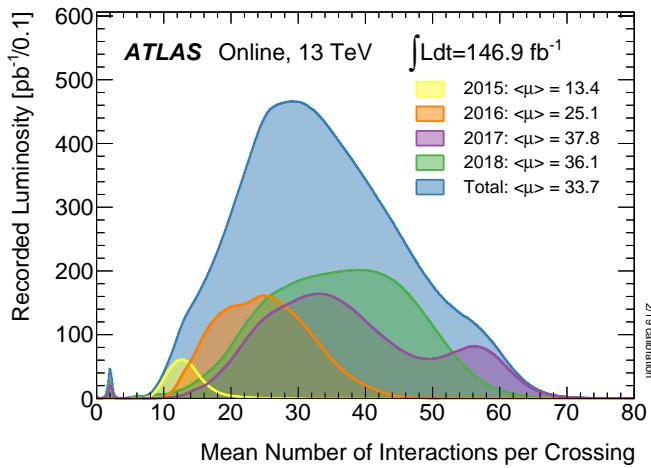
The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In total,  $139 \text{ fb}^{-1}$  of usable physics data was collected over the three-year run. The uncertainty on the total integrated luminosity is 1.7% [42] as obtained by the LUCID-2 detector [43].

## Pile-up

At the centre of the ATLAS detector, bunches of more than  $10^{11}$  protons are collided. Each bunch-crossing is called an *event*. There is generally at most one hard (i.e. high- $p_T$ ) proton-proton scatter per event. Additional interactions are typically relatively soft (low- $p_T$ ) and are known as *pile-up*. Pile-up from interactions within the same bunch-crossing is known as *in-time* pile-up while residual signatures from previous bunch-crossings is known as *out-of-time* pile-up. The number of pile-up interactions is denoted  $\mu$ , which is often given as a time-averaged value  $\langle \mu \rangle$ . Histograms showing the number of pile-up interactions over the course of Run 2 are shown in Fig. 3.5.



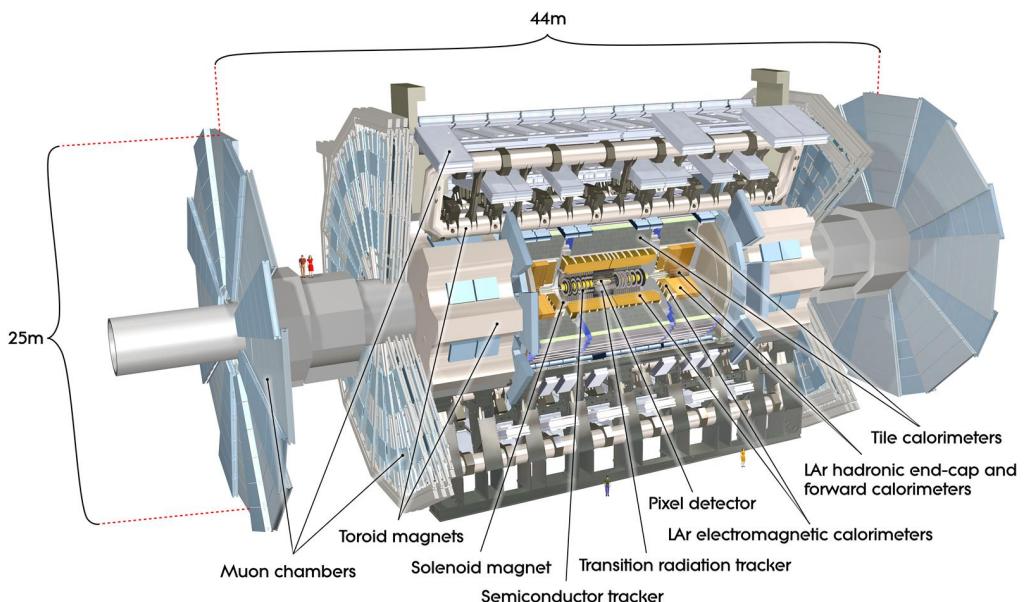
**Figure 3.4:** Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [37]. A total of  $139 \text{ fb}^{-1}$  of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally [44].



**Figure 3.5:** Average pile-up profiles measured by ATLAS during Run 2 [37].

### 3.3 The ATLAS Detector

The ATLAS<sup>2</sup> detector is made up of several specialised sub-detectors which are arranged concentrically around the nominal interaction point at the centre of the detector, as shown in Fig. 3.6. The detector is designed to cover nearly the entire solid angle around the collision point. In this section a brief overview of each sub-detector is given, in order of increasing radial distance from the point of collision. The inner tracking detector is described in Section 3.3.1, the electromagnetic and hadronic calorimeters in Section 3.3.2, and finally the muon spectrometer in Section 3.3.3. The trigger system, used to select events for storage, is described in Section 3.3.4. More complete information on the detector can be found in Ref. [35], while an overview of physics performance is given in [45].



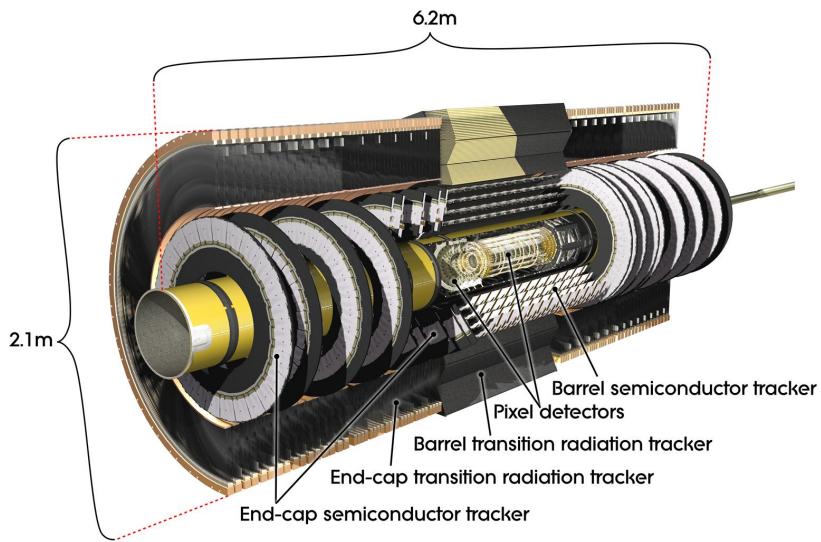
**Figure 3.6:** A 3D model of the entire ATLAS detector [46]. Cutouts to the centre of the detector reveal the different sub-detectors which are arranged in concentric layers around the nominal interaction point.

<sup>2</sup>A Toroidal LHC ApparatuS.

### 3.3.1 Inner Detector

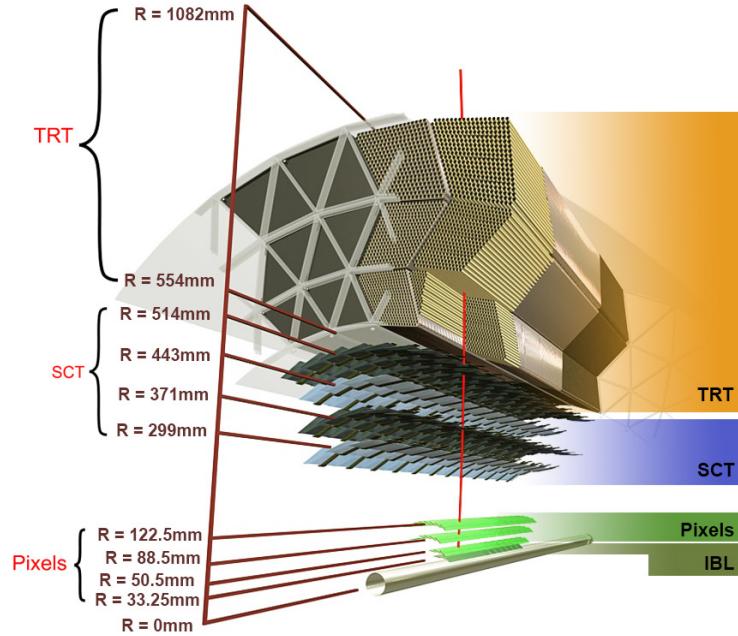
The inner-detector system (ID) provides high-resolution charged particle trajectory tracking in the range  $|\eta| < 2.5$ . Tracks are reconstructed from space-points obtained from clusters of charge depositions in the ID subsystems. The ID is immersed in a 2 T axial magnetic field, produced by a superconducting solenoidal magnet, which enables the measurement of particle momentum and charge. After Run 3, the ID will be replaced by the ITk [47, 48].

The inner detector is made up of several sub-systems, shown in Figs. 3.7 and 3.8. The high-granularity silicon pixel detector covers the innermost region and typically provides four space-point measurements per track. It is followed by the silicon microstrip tracker (SCT), which usually provides a further four space-point measurements (8 hits) per track. These silicon detectors are complemented by the Transition Radiation Tracker (TRT), which enables radially extended track reconstruction in the region  $|\eta| < 2.0$  and typically provides 33 (38) additional space-points in the barrel (end-cap).



**Figure 3.7:** A 3D model of the ATLAS ID showing the pixel, SCT and TRT sub-detectors [49].

The target inverse momentum resolution for the combined ID measurement is parameterised as a function of the track transverse momentum and polar angle [45].



**Figure 3.8:** A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [40].

The parameterisation is given by

$$\sigma(1/p_T) = 0.36 \oplus \frac{13}{p_T \sin \theta} \text{ TeV}^{-1}, \quad (3.7)$$

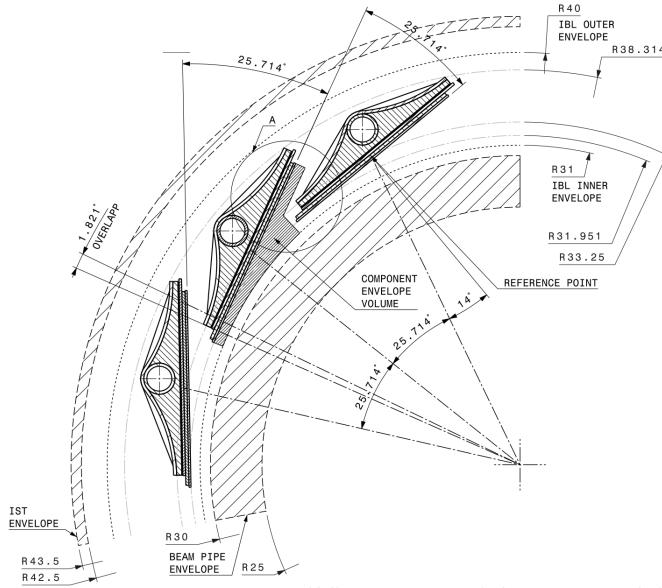
where  $\oplus$  denotes a sum in quadrature. For low- $p_T$  tracks (e.g.  $p_T \approx 500 \text{ MeV}$ ) in the central region this corresponds to a relative error of approximately 0.01%. Meanwhile for high- $p_T$  tracks (e.g.  $p_T \approx 100 \text{ GeV}$ ) in the central region this corresponds to a relative error of approximately 4%. The momentum resolution is generally good enough to correctly identify the sign of the charge on particles up to the highest energies observed at the LHC. The transverse impact parameter resolution  $\sigma(d_0)$  is parameterised similarly as

$$\sigma(d_0) = 11 \oplus \frac{115}{p_T \sin \theta} \mu\text{m}. \quad (3.8)$$

Typical uncertainties for the transverse IP resolution of tracks in the central region of the detector are 230  $\mu\text{m}$  and 11  $\mu\text{m}$  for low and high- $p_T$  tracks respectively.

# Pixel Detector

The silicon pixel detector is comprised of four cylindrical barrels at increasing radii from the beamline, and four disks on each side. The innermost barrel layer is the insertable B-layer (IBL), shown in Fig. 3.9. The IBL was installed before Run 2 [50,51] and lies approximately 33 mm from the beam axis (the distance varies due to the interleaved structure). The second-to-innermost layer is referred to as the B-layer. The specification of the pixel detector determines the impact parameter resolution and the ability to reconstruct primary and secondary vertices. The detector is required to have a high granularity (i.e. resolution) to maintain the low occupancy required to resolve the trajectories of nearby particles. Individual pixels are 50  $\mu\text{m}$  in the transverse direction  $R\phi$  and 400  $\mu\text{m}$  in the longitudinal  $z$  direction (250  $\mu\text{m}$  for the IBL). Cluster positions have a resolution of approximately 10  $\mu\text{m}$  in  $R\phi$  and 100  $\mu\text{m}$  in  $z$ .



**Figure 3.9:** A schematic cross-sectional view of the ATLAS IBL [50].

## Semi-Conductor Tracker (SCT)

The SCT is made up of four concentric barrel layers in the central region, and nine disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,

with a small stereo angle (20 mrad) between the two layers enabling the  $z$ -coordinate to be measured from a pair of strip measurements. The SCT typically provides four space-point measurements (eight strip measurements, or hits) per track in the barrel region. These have intrinsic uncertainties of 17  $\mu\text{m}$  in the transverse direction  $R\phi$ , and 580  $\mu\text{m}$  in the longitudinal direction  $z$  [52]. The measurements provide a contribution to the measurement of charged particle momentum and impact parameter.

### Transition Radiation Tracker (TRT)

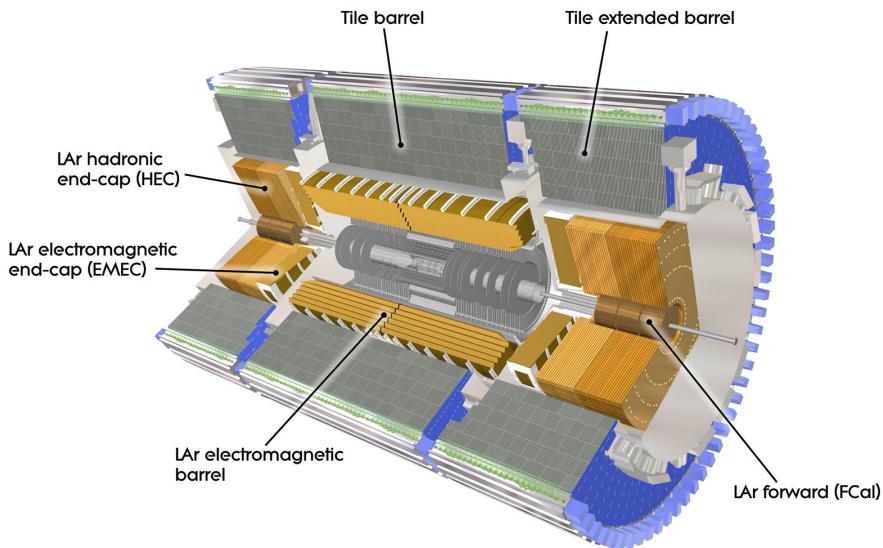
The TRT is a straw-tube tracker which complements the higher-resolution silicon-based tracks by offering a larger number of hits per track and a long lever arm, which aids the accurate measurement of particle momentum. It is made up of approximately 300 000 drift tubes with a diameter of 4 mm which are filled with an argon/xenon gas mixture. The walls of each tube are electrically charged, and a thin conducting wire runs along the center. When a charged particle traverses a tube, it ionises the gas and the resulting liberated electrons drift along the electric field to the wire, where an associated charge is registered. In the barrel the straws run parallel to the  $z$ -axis and therefore the TRT only provides tracking information in  $R\phi$ . Straws are arranged radially in the end-caps. The resulting two-dimensional space-points have a resolution of approximately 120  $\mu\text{m}$ . The spaces between the straws are filled with a polymer which encourages the emission of transition radiation, aiding electron identification.

### 3.3.2 Calorimeters

The calorimeter system measures the energy of incident particles in the range  $|\eta| < 4.9$ . There are three main sub-systems: the electromagnetic calorimeter (ECal), which focuses on the measurement of electrons and photons, and the hadronic calorimeter (HCal), which measures the energy of hadrons. The ECal and the HCal both instrument the region  $|\eta| < 3.2$ . In addition, a dedicated forward calorimeter subsystem (not described below) is used to measure the energy of particles in the range  $3.2 < |\eta| < 4.9$ .

A schematic view of the calorimeter system is shown in Fig. 3.10. Upon entering the calorimeter, incident particles will interact with the detector material to produce

a shower of secondary particles with reduced energies. The charge deposited in this process is measured and used to reconstruct the energy of the initial incident particle. The calorimeters must provide strong containment of showering particles to prevent punch-through of EM and hadronic particles to the HCal and muon systems respectively.



**Figure 3.10:** The ATLAS calorimeters [53]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic end-caps.

### Liquid Argon (LAr) Electromagnetic Calorimeter

The more granular lead/liquid-argon ECal covers the region  $|\eta| < 3.2$  and is split into barrel (covering  $|\eta| < 1.475$ ) and end-cap (covering  $1.375 < |\eta| < 3.2$ ) regions. EM calorimetry works by encouraging electrons and photons to interact with electrically charged particles in detector material via bremsstrahlung ( $e \rightarrow e\gamma$ ) and pair production ( $\gamma \rightarrow e^+e^-$ ). The EM calorimeter uses lead absorber plates to initiate EM showers, resulting in secondary particles which ionise the surrounding liquid argon. The charge is collected on copper electrodes and read out. The accordion geometry of the ECal allows for a full coverage in  $\phi$  without any azimuthal cracks.

The energy resolution of the LAr calorimeter is made up of a sampling, noise and constant term, which are summed in quadrature to produce the overall energy resolution, as in

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{120 \text{ MeV}}{E} \oplus 0.3\%. \quad (3.9)$$

The sampling term contributes approximately  $2.8\%/\sqrt{E}$ , the noise term contributes  $120 \text{ MeV}/E$ , while the constant term contributes an additional 0.3% and is dominant at high- $p_T$ . The resolution measurements are obtained from test beam data [54] and exceed the design specification [45]. The energy resolution for electrons with  $p_T = 10 \text{ GeV}$  varies between 0.3% and 0.8% over the pseudorapidity range. Meanwhile for photons with  $p_T = 60 \text{ GeV}$  the energy resolution varies between 0.25% and 1% [55].

### Hadronic Tile Calorimeter

In the central barrel region with  $|\eta| < 1.7$ , the HCal uses a tile calorimeter with steel as an absorbing material, and scintillating tiles as the active material. Two copper/liquid-argon calorimeter end-caps are also used which extend the coverage up to  $|\eta| = 3.2$ . Incident hadrons interact via the strong and electromagnetic forces with the absorber material, mainly loosing energy due to multiple inelastic nuclear collisions. The active material captures the resulting electrons and photons to measure the energy of the incident hadron.

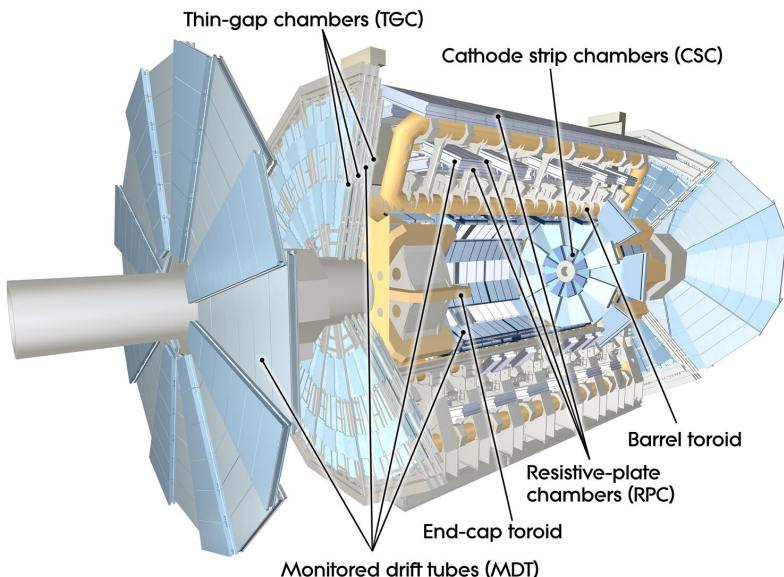
The granularity of the HCal is  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  in the region  $|\eta| < 2.5$ , and  $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$  in the region  $2.5 < |\eta| < 3.2$ . The hadronic energy resolution of the HCal is parameterised in the same way as the ECal. In a fit to test beam data [45], the energy resolution of single hadrons in the barrel is given by

$$\frac{\sigma(E)}{E} = \frac{52\%}{\sqrt{E}} \oplus \frac{1.6 \text{ GeV}}{E} \oplus 3\%. \quad (3.10)$$

This corresponds to an energy resolution of approximately 17% (6%) for a hadron with  $E_T$  of approximately 10 GeV (100 GeV).

### 3.3.3 Muon Spectrometer

Due to their higher mass, muons easily pass unimpeded through the ID and calorimeters and therefore require specialised detectors for their measurement. The Muon Spectrometer (MS) is made up of dedicated tracking and triggering hardware, as shown in Fig. 3.11. The precision tracking system uses three layers of monitored drift tubes with a barrel region covering  $|\eta| < 1.2$  and end-caps covering  $1 < |\eta| < 2.7$ . The inner layers of the end-caps use cathode strip chambers to better cope with the high occupancy in the forward region. The trigger system is comprised of resistive plate chambers in the barrel region covering  $|\eta| < 1.0$  and thin gap chambers in the end-cap regions covering  $1 < |\eta| < 2.4$ . A set of three superconducting air-core toroidal magnets, each made up of eight coils, is used in each of the barrel and end-caps to deflect the muons as they pass through the MS, allowing their momentum and charge to be measured from the direction and magnitude of curvature. The toroidal magnets generate a field which is largely orthogonal to the muon trajectories which allows for maximum deflection. The transverse momentum resolution (measured for combined ID and muon tracks, see Section 3.4.4) has been measured to be approximately 1.7% in the central region for muons from  $J/\psi$  decays, increasing to 2.3% for muons in the forward regions [56].



**Figure 3.11:** The ATLAS muon spectrometer [57].

### 3.3.4 The Trigger

The 25 ns bunch spacing used over the course of Run 2 corresponds to a bunch-crossing or event rate of 40 MHz (see Table 3.1). If the full information for the detector was written out for each event, this would correspond to the generation of 60 TB of data each second. This is more than can be feasibly processed and stored, requiring the use of a trigger system which quickly makes a decision about whether or not an event is potentially interesting and should be kept for further analysis. The trigger system is comprised of two levels which aim to identify various signatures, such as electrons, muons, taus, photons, and jets (including  $b$ -jets), as well as events with large total or missing transverse energy. The hardware-based Level-1 (L1) trigger uses coarse information from the calorimeters and MS to accept events at an average rate of 100 kHz approximately 2.5  $\mu$ s after the event. After the L1 trigger, the software-based High Level Trigger (HLT) makes use of 40 000 CPU cores to make a final selection on surviving events in approximately a few hundred milliseconds. The final event read-out rate is approximately 1.2 kHz, corresponding to  $1.2 \text{ GB s}^{-1}$  of permanent data storage. More information is provided in [58].

## 3.4 Reconstructed Physics Objects

Event reconstruction is the process of analysing the output from the detector to determine the type and properties of particles present in an event. The reconstructed event provides information about the underlying physics process that led to these observable final state particles. Events passing the trigger selection (described in Section 3.3.4) undergo offline reconstruction, which makes use of the full information from the detector. Reconstruction and analysis of events relies on the extensive ATLAS software stack, see Refs. [59, 60] for more information.

Several different reconstructed objects are used for physics analyses. Objects relevant to this thesis are described below.

### 3.4.1 Tracks

The reconstructed trajectories of charged particles are referred to as *tracks*. Tracks are reconstructed from the energy depositions (called *hits*) left by the particles as they traverse the inner detector. Tracks are used in the reconstruction of other objects, including vertices and jets, so their accurate reconstruction is a critical task. A comprehensive introduction to ATLAS tracking is available in Ref. [61], while specific optimisations for dense environments are detailed in Refs. [62, 63]. An overview of track reconstruction is given below.

#### Space-point Formation (Clustering)

When a charged particle traverses a silicon layer, charge can be collected in more than one pixel or strip. This is due to the incident angle of the particles with respect to the sensor, and also the drift of electrons between sensors caused by the magnetic field. Groups of neighbouring pixels or strips are clustered together and space-point coordinates are estimated using the shape and energy distribution of the clusters.

#### Track Finding

Space-points are used to build track seeds. These are groups of three hits which are geometrically compatible with being part of a track segment. A combinatorial Kalman filter (KF) is used to build *track candidates* by extending track seeds. The filter can create multiple track candidates per seed, with bifurcations along the track occurring when more than one compatible space-point exists on a given layer. In this way, the KF creates an excess of track candidates, which are only required to satisfy basic quality requirements. Track candidates are allowed to reuse or *share* hits freely (a single hit may be used by multiple track candidates). Typically, the presence of shared hits is a predictor of a bad track due to the high granularity of the ATLAS tracking detectors. At this stage, there can also be a large number of incorrect hits assigned to otherwise good tracks, and additionally large numbers of *fake* tracks, which are comprised of a majority of incorrect hits and do not correspond to the trajectory of any one physical particle (see Eq. (5.5)). The low quality of tracks at this stage necessitates an ambiguity solving step, in which candidates are cleaned, and the highest quality tracks are selected.

## Ambiguity Solving

Ambiguity solving was introduced as part of the ATLAS New Tracking effort [61], which was intended to improve track reconstruction performance in dense environments. In the ambiguity solver, track candidates are processed individually in descending order of a track score. The track score quantifies the likelihood of the track corresponding to the trajectory of a real particle. Scoring uses a number of variables, including the number and positions of hits (preferring hits in more precise regions of the detector), the transverse momentum of the track and the track fit quality. The track fit quality is defined as the  $\chi^2$  obtained from the track fit divided by the number degrees of freedom on the track. A preference for high transverse momentum tracks promotes the successful reconstruction of the more physically interesting energetic particles, and suppresses the large number of incorrect hits assigned to low momentum tracks. The ambiguity solver also penalises tracks with missing hits on the innermost detector layers.

During the processing of a track candidate, the track is cleaned (whereby problematic hits are removed), and, if the resulting track satisfies the quality selection criteria, a high precision fit of the track parameters using the surviving hits is performed. The high precision fit makes full use of all available information, and uses an updated position and uncertainty estimate for each cluster obtained from a Neural Network (NN) [64]. If the track has reached this stage without being rejected by passing various quality requirements, it is re-scored and returned to the list of track candidates. If the same track is then processed again without requiring modification, it is added to the final track collection. Track candidates that fall below certain quality threshold are rejected. This selection does allow for the possibility of a track having small number of shared hits, as detailed in Table 3.2.

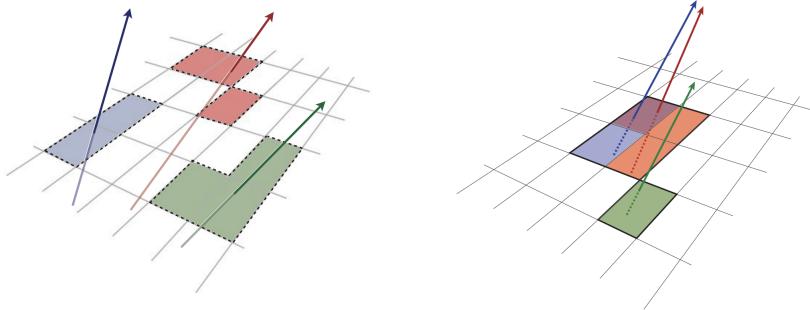
## Neural Network Cluster Splitting

As part of track cleaning, shared hits are classified by a NN to determine if they are compatible with the characteristic features of a merged cluster [62, 64]. A merged cluster is one made up of a combination of energy deposits from more than one particle which have become merged due to the closeness of the particles and the limited resolution of the detector. It is common for clusters to become merged in

Parameter	Selection
$p_T$	$> 500 \text{ MeV}$
$ \eta $	$< 2.5$
$ d_0 $	$< 3.5 \text{ mm}$
$ z_0 \sin \theta $	$< 5 \text{ mm}$
Silicon hits	$\geq 8$
Shared silicon hits	$< 2$
Silicon holes	$< 3$
Pixel holes	$< 2$

**Table 3.2:** Quality selections applied to tracks, where  $d_0$  is the transverse IP of the track,  $z_0$  is the longitudinal IP with respect to the primary vertex and  $\theta$  is the track polar angle (see Section 3.2.2 for the IP definitions). Silicon hits are hits on the pixel and SCT layers. Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [63]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

dense environments, as discussed in Section 4.2. If the cluster is predicted to be merged it is labelled as being freely shareable, or *split*. Hits not compatible with the merged hypothesis can still be shared by a limited number of tracks, but come with a penalty for the track which may hinder its acceptance into the final track collection.



**Figure 3.12:** (left) Particles which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [63].

## Pseudotracking

In simulation, pseudotracking uses Monte Carlo truth information to group together all the hits left by each truth particle. Each collection of hits which, as a unit, satisfies basic quality requirements is directly used in a full resolution track fit. If the track fit is successful, a “pseudotrack” track is created and stored. If the track fit fails, or the collection of hits does not pass the basic quality requirements (for example because of a lack of hits) then the particle is said to be un-reconstructable. In this way, pseudotracking performance represents the ideal reconstruction performance given the ATLAS detector, with perfect hit-to-track association and track reconstruction efficiency. The approach was introduced in Ref. [65] as a way to obtain a fast approximation of tracking reconstruction for simulated data, however the technique has become a useful tool for studying tracking performance in general [62].

### 3.4.2 Vertices

Groups of reconstructed tracks can be examined to determine whether the particles originated from a common spatial point of origin. This occurs when proton-proton collisions take place (primary vertices), when a particle decays or radiates, and also as a result of interaction with the detector material (secondary vertices). Vertex reconstruction is made up of two stages. First, vertex finding takes place, which is the process of grouping tracks into compatible vertices. Second, vertex fitting combines information from compatible tracks to reconstruct the physical properties of the vertex, such as mass and position.

#### Primary Vertices

Each proton-proton interaction happens at a *primary vertex*. Primary vertices are iteratively reconstructed with tracks using the iterative vertex finder [66]. The *hard scatter primary vertex* is defined as the primary vertex whose associated tracks have the largest sum of transverse momentum squared,  $\Sigma(p_T^2)$  [67]. In the following, the “primary vertex” always refers to the hard scatter primary vertex.

## Secondary Vertices

Secondary vertices (SV) occur when a particle radiates or decays at a sufficient distance from the primary vertex to be resolved from the primary vertex (see Section 3.5.2). Two widely used secondary vertexing tools are used within ATLAS: SV1 and JetFitter [68, 69]. Each attempts to reconstruct secondary vertices inside a jet using the tracks associated to that jet (see Section 3.4.3 for more information about jet reconstruction and track association). SV1 by design attempts to reconstruct only a single inclusive vertex per jet. This inclusive vertex groups all  $b$ -hadron decay products, including tracks from the  $b$ -hadron decay itself and tracks from  $b \rightarrow c$  decays. The second tool, JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary vertices from  $b$ -hadron decays are reconstructed separately to tertiary vertices from  $b \rightarrow c$  decay chains.

### 3.4.3 Jets

Jets are an aggregate reconstructed object corresponding to a collection of collimated stable particles which results from the presence of a quark or gluon. Jets are built by clustering constituent objects (e.g. tracks or calorimeter clusters) using a jet finding algorithm, for example the anti- $k_t$  algorithm [70], which is implemented in FASTJET [71].

Objects can be associated to jets in one of two ways. The first is via a geometrical matching in  $\Delta R$  (see Eq. (3.3)). The second is via a ghost association [72], where the object is assigned a negligible momentum and re-clustered into the jet after its formation.

Jets from pile-up interactions are suppressed using the Jet Vertex Tagger (JVT) algorithm, which uses the transverse momenta of tracks to identify jets from pile-up interactions [73].

Different types of jets used in this thesis are described below.

## Particle Flow Jets

Particle-flow (PFlow) jets are reconstructed from particle-flow objects [74] using the anti- $k_t$  algorithm with a radius parameter of 0.4. Particle-flow objects integrate information from both the ID and the calorimeters. The addition of tracking information from the ID results in improvements to the energy resolution at low transverse momenta and improvements to the position resolution at high transverse momenta. Additionally tracking information can be used to reduce pile-up contamination. The PFlow jet energy scale is calibrated according to Ref. [75]. PFlow jets are used to train the algorithms discussed in Chapter 5 and Chapter 6.

## EMTopo Jets

EMTopo jets are reconstructed from noise-suppressed topological clusters (topoclusters) of calorimeter energy depositions [76]. The clustering uses the energy significance of each cell, defined as

$$S_{\text{cell}} = \frac{E_{\text{cell}}}{\sigma_{\text{noise, cell}}}, \quad (3.11)$$

where  $E_{\text{cell}}$  is the energy measured in a given calorimeter cell, and  $\sigma_{\text{noise, cell}}$  is the expected level of noise on the cell (e.g. from pile-up interactions). Topoclusters are formed from a seed cell with a large  $S_{\text{cell}}$ , and expanded by iteratively adding neighbouring cells with a sufficiently large energy significance. Collections of topoclusters are then clustered into a jet using the anti- $k_t$  algorithm with a radius parameter of 0.4 (small- $R$  jets) or 1.0 (large- $R$  jets). More information, including information on the calibration of the topocluster jet energy scale, is available in Ref. [76]. EMTopo jets are used in the analysis described in Chapter 7.

## Large- $R$ Jets

Large- $R$  jets have a radius parameter  $R = 1.0$  and are built by clustering topological calorimeter clusters using the anti- $k_t$  algorithm [77]. The large radius parameter is especially useful for containing the decay products of a boosted Higgs boson, as discussed in Chapter 7. Due to their large size, large- $R$  jets benefit from a grooming procedure called trimming which remove soft contaminants inside the jet [78, 79].

Trimming aims to remove jet constituents from pile-up and the underlying event, which helps to improve the jet mass resolution and its robustness to varying levels of pile-up [80]. The jet mass is computed using a combination of information from the calorimeters and ID, and a calibration to data is applied [81]. Large- $R$  jets are used in the analysis described in Chapter 7.

### Track-jets

Variable-radius track-jets are built by clustering tracks using the anti- $k_t$  clustering algorithm. They are associated to large- $R$  jets as sub-jets and used to identify large- $R$  jets containing  $b$ -hadrons. The radius parameter is allowed to vary with transverse momentum such that a broader cone (up to  $R = 0.4$ ) is used for low- $p_T$  track-jets and a narrower cone (down to  $R = 0.02$ ) for high- $p_T$  track-jets [82, 83]. The narrower cone is better suited to clustering highly collimated jet constituents at high- $p_T$ . Track-jets are used in the analysis described in Chapter 7.

### Jet Flavour Labels

In simulation, jet flavour labels are assigned to small- $R$  jets according to the presence of a generator level hadron within  $\Delta R(\text{hadron}, \text{jet}) < 0.3$  of the jet axis. If a  $b$ -hadron is found the jet is labelled a  $b$ -jet. In the absence of a  $b$ -hadron, if a  $c$ -hadron is found the jet is called a  $c$ -jet. If no  $b$ - or  $c$ -hadrons are found, but a  $\tau$  is found in the jet, it is labelled as a  $\tau$ -jet. Otherwise, the jet is labelled as a light-jet.

### Jet-to-Track Association

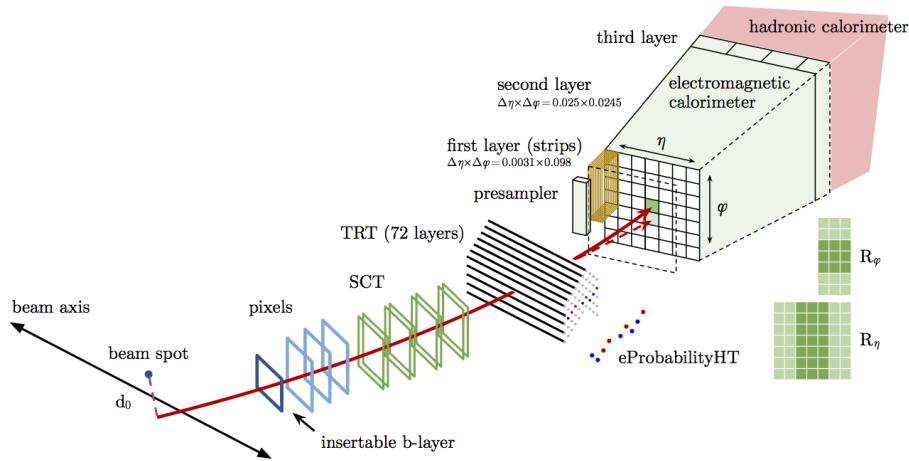
For the purposes of flavour tagging (see Section 3.5), tracks are associated to small- $R$  jets using a  $\Delta R$  association cone, the width of which decreases as a function of jet  $p_T$ , with a maximum cone size of  $\Delta R \approx 0.45$  for jets with  $p_T = 20\,\text{GeV}$  and minimum cone size of  $\Delta R \approx 0.25$  for jets with  $p_T > 200\,\text{GeV}$ . If a track is within the association cones of more than one jet, it is assigned to the jet which has a smaller  $\Delta R(\text{track}, \text{jet})$ .

### 3.4.4 Leptons

Electrons and muons leave characteristic signatures that are picked up in the ECal and MS respectively. The reconstruction of both types of charged lepton is briefly outlined below.

#### Electrons

A diagrammatic view of electron reconstruction is shown in Fig. 3.13. Electron candidates are reconstructed by matching primary vertex compatible<sup>3</sup> inner detector tracks to topological calorimeter clusters. The track-cluster matching criteria takes into account the significant energy loss of the electron due to bremsstrahlung. If a match is found, a refit of the track is performed using the Gaussian Sum Filter (GSF) [84], which better handles trajectory reconstruction in the presence of bremsstrahlung. Various identification criteria are then applied to the candidates using a likelihood-based (LH) method to improve purity. These include requirements on the track quality and cluster matching, the shape of electromagnetic shower in the ECal, leakage into the HCal, and the amount of transition radiation detected in the TRT. Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may also be applied. A more complete description can be found in Refs. [85, 86].



**Figure 3.13:** A sketch of electron reconstruction using the ATLAS detector [87]. Electron reconstruction makes use of the entire ID and the calorimeters.

<sup>3</sup>The ID track associated with the electron is required to satisfy  $d_0/s(d_0) < 5$  and  $z_0 \sin \theta < 0.5$  mm.

## Muons

Muon reconstruction makes use of the dedicated MS (see Section 3.3.3), the tracks from the ID, and the presence of characteristic signatures in the calorimeters. Muon tracks (i.e. tracks reconstructed in the MS) are reconstructed by connecting straight-line track segments, which are identified via a Hough transform, and combined into an approximately parabolic trajectory. Finally, a global  $\chi^2$  fit is performed, taking into account possible interactions between the muon and the detector material. A reconstructed muon is called *combined* if it can be matched successfully to an ID track. Combined muons undergo a further fit with the combined ID and MS hits, with the energy loss due to the traversal of the calorimeters being taken into account.

After reconstruction, muon candidates undergo an identification processes which helps to efficiently identify prompt muons whilst rejecting background signals (e.g. non-prompt muons from pion, kaon, semi-leptonic heavy flavour hadron decays, and the punch-through of a hadron from the calorimeter). Combined muon identification takes into account discrepancies in the  $p_T$  and charge measurements between the MS and ID and the  $\chi^2$  of the combined track fit. Selections on the number of hits in the ID and MS are also applied. At the medium identification working point, approximately 96% of prompt muons with  $20 \text{ GeV} < p_T < 100 \text{ GeV}$  are successfully identified. On top of the identification requirements, a number of isolation requirements can also be applied to further suppress background signals.

More information on muon reconstruction, identification and isolation can be found in Ref. [56].

### 3.4.5 Missing Transverse Momentum

An imbalance in the final state transverse momentum can occur as a result of incomplete measurement of the final state particles. In particular, neutrinos are not measured by the detector and contribute to the missing transverse momentum  $\mathbf{E}_T^{\text{miss}}$ . Incomplete detector acceptance and inaccuracies in the reconstruction of the final state can also contribute to the missing transverse momentum in an event. In order to calculate the missing transverse momentum, the negative vector sum of the momentum of all photons, leptons and small- $R$  jets with  $p_T > 20 \text{ GeV}$  is taken.

The momenta of tracks associated to the primary vertex are also taken into account. The magnitude of  $\mathbf{E}_T^{\text{miss}}$  is written  $E_T^{\text{miss}}$ . More information about missing transverse momentum reconstruction is provided in [88].

## 3.5 Flavour Tagging Algorithms

Many ATLAS analyses rely on flavour tagging, which is the identification of jets containing heavy flavour hadrons ( $b$ -hadrons and  $c$ -hadrons) as opposed to those containing only light-flavour hadrons or gluons (collectively referred to as light-jets). In particular,  $b$ -tagging is the identification of jets originating only from  $b$ -hadrons (i.e.  $b$ -jets). Flavour tagging is a critical component of the physics programme of the ATLAS experiment. It is of crucial importance for the study of the Standard Model (SM) Higgs boson and the top quark, which decay preferentially to  $b$ -quarks [89, 90], and additionally for several Beyond the Standard Model (BSM) resonances that readily decay to heavy flavour quarks [91].

In this section some common definitions used in flavour tagging are given in Section 3.5.1, followed by a description of the characteristic properties of  $b$ -jets in Section 3.5.2. The various  $b$ -jet identification algorithms (also called *taggers*) used in ATLAS are introduced in Sections 3.5.3 and 3.5.4.

### 3.5.1 Common Definitions

Each flavour tagging algorithm outputs a discriminant value that can be used to select jets of the signal flavour. The efficiency of a flavour tagging algorithm is defined as the fraction of signal jets which are correctly identified by the tagger. Given  $N^{\text{signal}}$  signal jets,  $N_{\text{tagged}}^{\text{signal}}$  of which have passed the algorithm’s selection and  $N_{\text{untagged}}^{\text{signal}}$  of which have not, the efficiency is given by

$$\varepsilon = \frac{N_{\text{tagged}}^{\text{signal}}}{N_{\text{tagged}}^{\text{signal}} + N_{\text{untagged}}^{\text{signal}}}. \quad (3.12)$$

Meanwhile the *fake rate* is defined as the efficiency for a background class to be selected, i.e.

$$\text{fake rate} = \frac{N_{\text{tagged}}^{\text{background}}}{N_{\text{tagged}}^{\text{background}} + N_{\text{untagged}}^{\text{background}}}. \quad (3.13)$$

The *rejection power* of the model is quantified for a given background as the reciprocal of the fake rate, i.e.

$$\text{rejection power} = \frac{1}{\text{fake rate}}. \quad (3.14)$$

A *fixed-cut working point* (WP) defines the corresponding selection applied to the tagging discriminant in order to achieve a given signal identification efficiency.

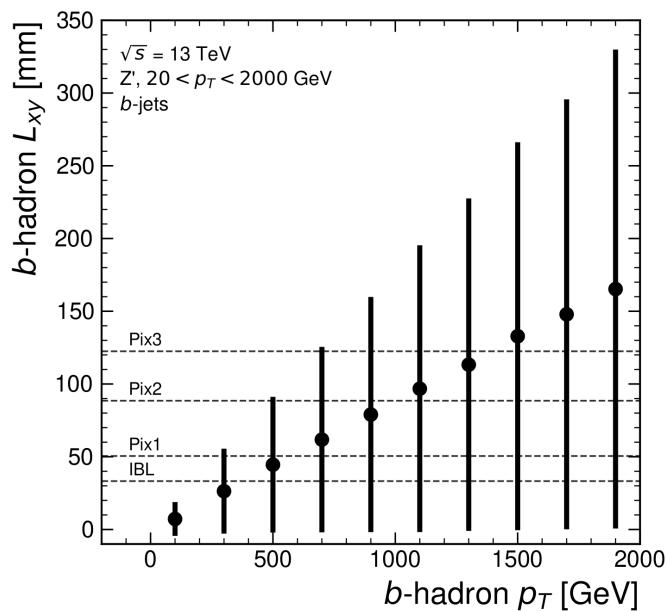
### 3.5.2 Decay Topology

$b$ -hadrons are quasi-stable bound states of a bottom quark and one or more lighter quarks. Collectively, these are the  $B$ -mesons (e.g.  $B^+ = u\bar{b}$ ,  $B^0 = d\bar{b}$ ) and baryons (e.g.  $\Lambda_b^0 = udb$ ). After a  $b$ -quark is produced as the result of a proton-proton collision, they quickly hadronise. The hadronisation process is hard – around 70-80% of the  $b$ -quark’s momentum is passed to the  $b$ -hadron, with the rest being radiated as prompt hadronisation or fragmentation particles. See Ref. [92] for a more in depth discussion on hadronisation and the closely related process of fragmentation. Henceforth the combined hadronisation and fragmentation products will be referred to collectively as fragmentation.

$b$ -hadrons are interesting objects of study due to their relatively long proper lifetimes  $\tau \approx 1.5 \text{ ps}$  [93]. This lifetime corresponds to a proper decay length  $c\tau \approx 450 \mu\text{m}$ . In the rest frame of the detector, the typical  $b$ -hadron travels a distance

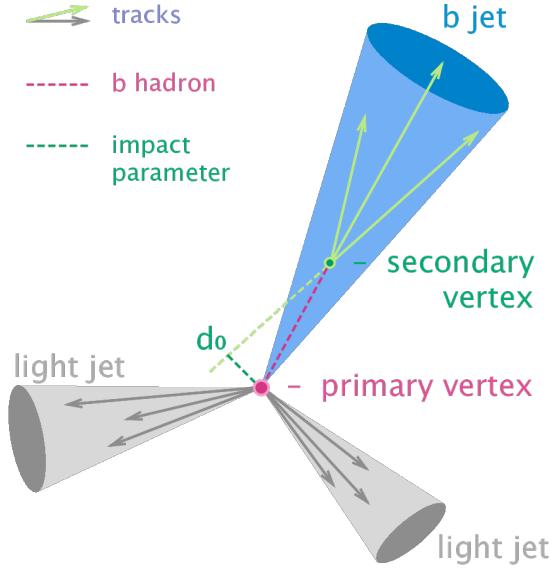
$$d = \gamma\beta c\tau \approx \gamma c\tau \quad (3.15)$$

before decaying, where in the high energy limit  $\gamma = E_b/m_b$  and  $\beta = v/c = 1$ . For a 50 GeV  $b$ -hadron, this gives  $d \approx 4.5$  mm, which is displaced enough to be resolved from the primary vertex. Meanwhile for a 1 TeV  $b$ -hadron  $d \approx 90$  mm – well beyond the radius of the first pixel layer (the IBL) which is situated at a radius of approximately 33 mm from the center of the detector. Fig. 3.14 shows how the mean decay radius varies as a function of  $b$ -hadron  $p_T$ . This significant displacement is characteristic of  $b$ -jets and makes it possible to reconstruct secondary vertices (SV) at the  $b$ -hadron decay point.



**Figure 3.14:** The truth  $b$ -hadron decay radius  $L_{xy}$  as a function of truth transverse momentum  $p_T$  for reconstructed  $b$ -jets in  $Z'$  events (see Section 4.1). Error bars show the standard deviation of the  $L_{xy}$  distribution in each  $p_T$  bin. The pixel layers are shown in dashed horizontal lines.

$b$ -hadrons decay weakly to on average four or five collimated stable particles [94]. These particles, along with any other fragmentation particles, are reconstructed in the detector as a jet. A  $b$ -jet has several characteristic features which differentiate it from light-jets. The primary feature is the presence of a high mass secondary vertex that is significantly displaced from the primary vertex. Reconstruction of these vertices from tracks with common points of spatial origin is a common approach used in the identification of  $b$ -jets.



**Figure 3.15:** Diagram of a typical  $b$ -jet (blue) which has been produced in an event alongside two light jets (grey) [95]. The  $b$ -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter  $d_0$  is a characteristic property of the trajectories of  $b$ -hadron decay products.

Additional signatures of  $b$ -hadrons are as follows. Associated tracks and SVs can have a large transverse impact parameter  $d_0$  as a result of the  $b$ -hadron displacement (as shown in Fig. 3.15). Furthermore, since it is common for the  $b$ -hadron to decay to a  $c$ -hadron with non-negligible lifetime, tertiary vertices can be found within  $b$ -jets resulting from  $b \rightarrow c$  decay chains. Additionally, as the  $b$ - or  $c$ -hadron decays semileptonically in approximately 40% of cases [18] the presence of a reconstructed electron or muon inside a jet provides information about the jet flavour.

These signatures are primarily identified using tracks associated to jets, or using reconstructed electrons or muons, which also rely on tracks as discussed in Section 3.4.4. As such, efficient and accurate track reconstruction is essential for high performance flavour tagging.

### 3.5.3 Low-level Algorithms

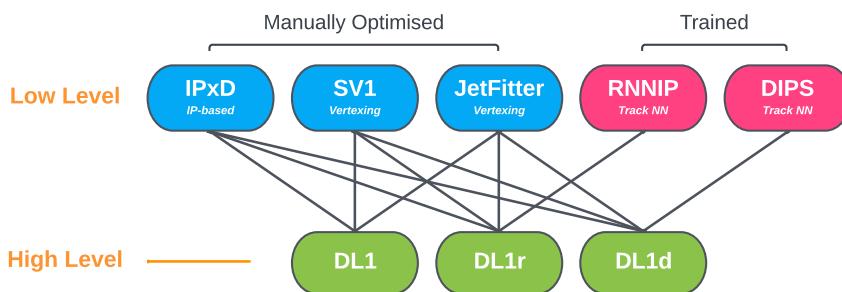
The low-level algorithms are the first step in the  $b$ -tagging process. They take inputs information about the jet and its associated tracks (see Section 3.4.3) and

reconstruct distinct features of the experimental signature of heavy flavour jets. The low-level algorithms are a combination of manually optimised reconstruction algorithms, for example SV1 and JetFitter which reconstruct displaced decay vertices (see Section 3.4.2) and IPxD which discriminates based on track IPs, and trained machine learning models such as RNNIP [96] and DIPS [97] which use IP and hit information from a variable number of tracks to predict the jet flavour.

### 3.5.4 High-level Algorithms

The high-level algorithms combine the outputs of the independently optimised low-level algorithms using a multivariate approach to produce a discriminant value for each jet. For example the MV2c10 algorithm [68,98,99], used in the analysis described in Chapter 7, consists of a boosted decision tree which combines the outputs of IPxD, SV1 and JetFitter. The working point is tuned to achieve an average  $b$ -jet efficiency of 70% on simulated  $t\bar{t}$  events. At this efficiency working point, rejection factors for  $c$ -jets and light-jets are approximately 9 and 304 respectively.

The current ATLAS flavour tagger, DL1r [99], is part of a series of taggers that use a deep neural network to combine the outputs of the low-level algorithms. At the 70% efficiency working point, DL1r achieves a  $c$ -jet rejection of approximately 12 and a light-jet rejection of approximately 625. The low-level algorithms used by the different high-level taggers DL1, DL1r and DL1d is shown in Fig. 3.16.



**Figure 3.16:** An overview of different low and high level taggers used in ATLAS. The low level taggers are IPxD, SV1 and JetFitter, and RNNIP and DIPS [68,69,96,97]. The outputs of these taggers are fed into the high-level taggers DL1, DL1r and DL1d [99,100].

# Chapter 4

## Tracking and Flavour Tagging at High- $p_{\mathrm{T}}$

The various flavour tagging algorithms introduced in Section 3.5 work by identifying the unique signatures of heavy flavour jets ( $b$ - and  $c$ -jets). Ultimately, the tagging algorithms use input information about the reconstructed jet and its associated tracks. Successful  $b$ -tagging therefore relies on the efficient and accurate track reconstruction, especially for tracks corresponding to the products of heavy flavour decays. In this chapter the challenges facing track reconstruction and flavour tagging at high- $p_{\mathrm{T}}$  are discussed.

The chapter is structured as follows. In Section 4.1 an introduction to the datasets used in this thesis is given. A summary of the challenges facing tracking and  $b$ -tagging at high transverse momentum is provided in Section 4.2. Some preliminary investigations into improving tracking in the high- $p_{\mathrm{T}}$  regime are investigated in Section 4.3. Finally, in Section 4.4 the conclusions of the chapter are given.

### 4.1 Datasets

This thesis makes extensive use of two simulated datasets which are described in this section. The datasets are made up of simulated SM  $t\bar{t}$  and BSM  $Z'$  events<sup>1</sup> initiated by proton-proton collisions at a center of mass energy  $\sqrt{s} = 13 \text{ TeV}$ . While the  $t\bar{t}$  sample populates the low- $p_{\mathrm{T}}$  phase space, the  $Z'$  sample is constructed in

---

<sup>1</sup>The  $Z'$  boson used in this thesis is a modified SM  $Z$  with an increased mass

such a manner that it has a relatively flat jet  $p_{\mathrm{T}}$  spectrum up to 5 TeV and decays democratically to equal numbers of  $b$ -,  $c$ - and light-jets. As a result, the  $Z'$  sample is well suited to the study of  $b$ -tagging at high- $p_{\mathrm{T}}$ . The simulation includes the effect of multiple proton-proton interactions per bunch crossing with an average pile-up of  $\langle \mu \rangle = 40$ . The effect on the detector response due to interactions from bunch crossings before or after the one containing the hard interaction are also included.

For the  $t\bar{t}$  sample, events are generated using the POWHEGBOX v2 generator [101–104] at next-to-leading order in the strong coupling constant  $\alpha_s$ . The NNPDF3.0NNLO [105] set of parton distribution functions (PDFs) are used for the calculation of the hard scatter matrix element. The  $h_{\mathrm{damp}}$  parameter<sup>2</sup> is set to 1.5 times the mass of the top-quark [106], with  $m_{\mathrm{top}} = 172.5$  GeV. The simulated hard scatter events are interfaced with PYTHIA 8.230 [107] using the A14 parameter tune and the NNPDF2.3LO PDFs to handle the simulation of the parton shower, hadronisation, and underlying event. These choices were found to best model the top quark transverse momentum and the number of additional jets in the event [106, 108]. Meanwhile for the  $Z'$  sample, full events are generated with PYTHIA 8.212. Again, the A14 tune [109] and the NNPDF2.3LO set of PDFs [110] are used.

For both samples the decays of  $b$ - and  $c$ -hadrons are performed by EVTGEN v1.6.0 [111]. After event generation, simulated particles are passed through the full ATLAS detector simulation [112] which is based on GEANT4 [113]. The interaction between the long-flying heavy flavour hadrons and the detector material is included in the simulation.

Additional jet requirements are as follows. Jets are required to have a pseudorapidity  $|\eta| < 2.5$  and  $p_{\mathrm{T}} > 20$  GeV. Jets are also required not to overlap with a prompt generator-level electron or muon from  $W$  boson decays. Finally, a standard selection using the JVT tagger (see Section 3.4.3) at the tight working point is applied to jets with  $p_{\mathrm{T}} < 60$  GeV and  $|\eta| < 2.4$  in order to suppress pile-up contamination [73].

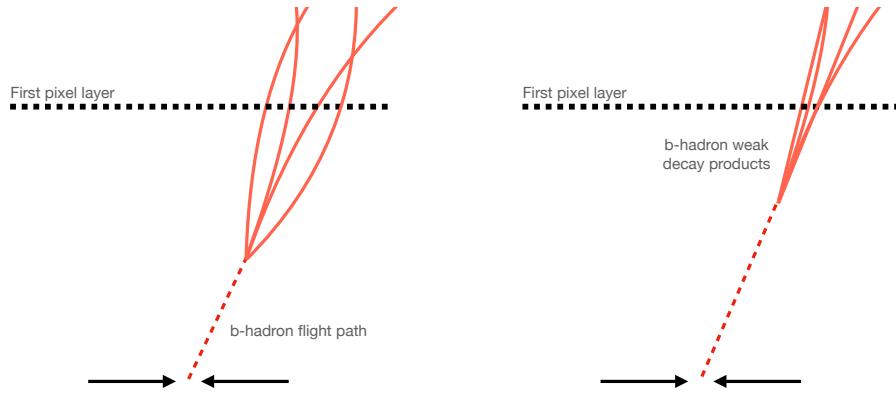
---

<sup>2</sup>The first gluon emission cut-off scale parameter  $h_{\mathrm{damp}}$  of the POWHEGBOX generator is used to limit the effect of resummed higher order corrections. It is used to suppress the transverse momentum of the radiation which the  $t\bar{t}$  system recoils against.

## 4.2 $b$ -hadron Reconstruction Challenges

As discussed in Section 3.5.2, a necessary requirement for successful  $b$ -tagging is the efficient and accurate reconstruction of the charged particle trajectories in the jet. For high  $p_T$  jets ( $p_T > 250$  GeV) this task becomes difficult due to difficulties in the accurate reconstruction of tracks, as described below.

As the  $b$ -jet energy increases, the multiplicity of the fragmentation products inside the jet increases, while the multiplicity of the products of the weak decay remains fixed. The “signal” tracks (those from the weak decay of the  $b$ -hadron) therefore become significantly outnumbered. Both fragmentation and  $b$ -hadron weak decay products also become increasingly collimated as their inherited transverse momentum increases. This is compounded by the increased decay length of  $b$ -hadrons (and  $c$ -hadrons) at high- $p_T$ , which means that the decay products have less of an opportunity to diverge before reaching the first tracking layers of the detector (shown in Fig. 4.1). If the weak decay of the  $b$ -hadron takes place close enough to a detector layer, or if the particles are otherwise sufficiently collimated, charge deposits left by nearby particles may not be resolved individually, instead being reconstructed as merged clusters [63].



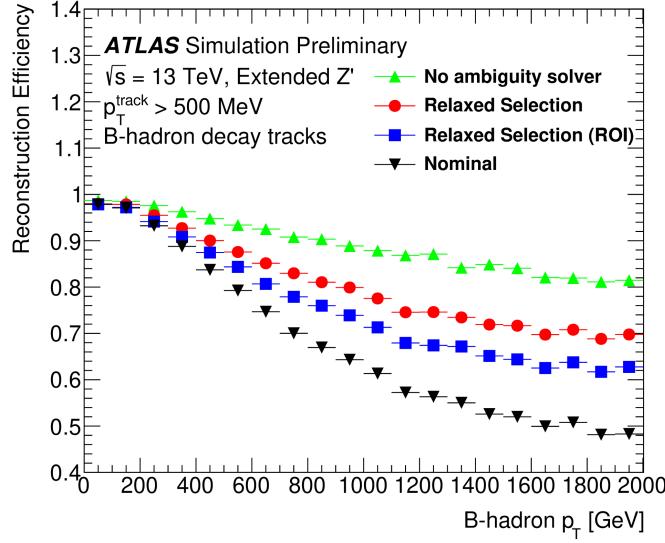
**Figure 4.1:** At lower  $p_T$  (left) the decay length of the  $b$ -hadron is on average reduced, and the decay tracks are less collimated. At higher  $p_T$  (right) the  $b$ -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching detector elements. As a result, the ID may be unable to resolve charged depositions from different particles, resulting merged clusters.

As discussed in Section 3.4.1, merged clusters are generally rare, and so shared hits generally predict bad tracks and are correspondingly penalised during track

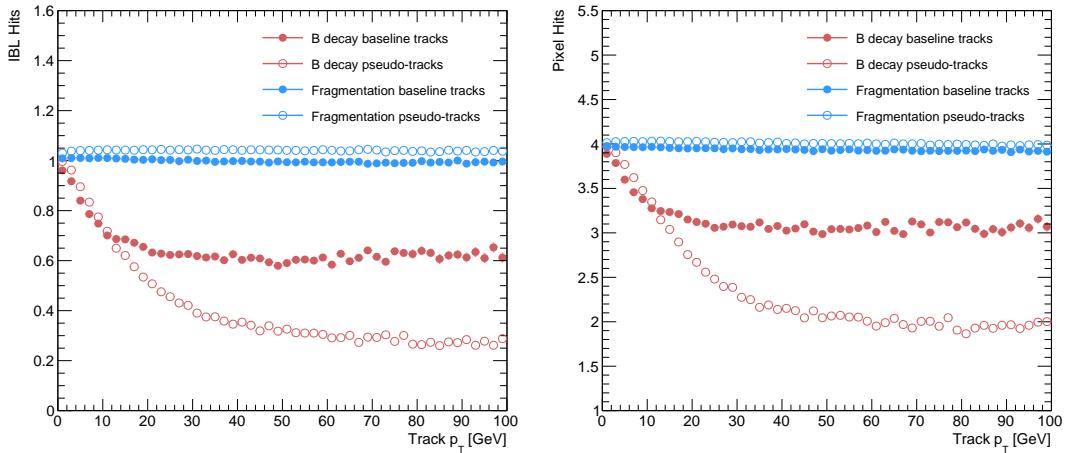
reconstruction. However, in the core of high  $p_{\mathrm{T}}$   $b$ -jets the density of particles is high enough that the probability of cluster merging increases dramatically. Successful reconstruction of such tracks requires the presence of shared hits to be effectively dealt with, but in the standard reconstruction the presence of these can instead impair the successfully reconstruction of the track. Furthermore, heavy flavour decays may also take place inside the tracking detectors themselves, which at best leads to missing measurements on the most sensitive detector layers, and at worst can lead to wrong inner layer hits being added to displaced tracks, since the reconstruction process penalises tracks without inner layer hits.

The above effects create two distinct but related problems for  $b$ -tagging. The first is a drop in track reconstruction efficiency. The presence of shared and missing hits reduces a track's score in the ambiguity solver meaning that higher ranking, but potentially less accurate, track candidates are processed first and take ownership of the hits. This can make it difficult for otherwise reasonable  $b$ -hadron decay tracks to meet the ambiguity solver's stringent track quality requirements, leading to their rejection at this stage and an overall decrease in the  $b$ -hadron decay track reconstruction efficiency. As shown in Fig. 4.2, this can result in a large drop in reconstruction efficiency for  $b$ -hadron decay products of up to 50% for at  $p_{\mathrm{T}} > 2 \text{ TeV}$ .

The second part of the problem is that, due to the high multiplicity of clusters available for assignment in the vicinity of the typical high- $p_{\mathrm{T}}$   $b$ -hadron decay track, and also given the strong positive bias of the ambiguity solver towards those tracks with pixel measurements in each layer (especially the innermost IBL measurement), many  $b$ -hadron decay tracks are assigned incorrect inner layer hits. This is only a problem for those decay products which were produced within the pixel detector as a result of a significantly displaced  $b$ -hadron decay, and so do not have a correct hit available for assignment. Fig. 4.3 shows the number of hits as a function of the reconstructed track  $p_{\mathrm{T}}$  for fragmentation tracks and tracks from the weak decay of the  $b$ -hadron. The baseline tracks represent the standard reconstruction setup, while the pseudotracks represent the ideal tracking setup as outlined in Section 3.4.1. Hit multiplicities on the pseudotracks decrease with increasing  $p_{\mathrm{T}}$  due to the flight of the  $b$ -hadron before its decay. The baseline tracks have more hits than the pseudotracks, indicating that they are being incorrectly assigned additional hits on the inner layers of the detector.



**Figure 4.2:**  $b$ -hadron decay track reconstruction efficiency as a function of truth  $b$ -hadron  $p_T$  for jets in the  $Z'$  sample [114]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- $p_T$   $b$ -hadrons, the ambiguity solver is overly aggressive in its removal of  $b$ -hadron decay tracks. Suggestions for the improvement of the track reconstruction efficiency in this regime by the loosening of cuts in the ambiguity solver are shown in blue and red.



**Figure 4.3:** Average hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the  $p_T$  of the reconstructed track for tracks in jets in a  $Z'$  sample at  $\sqrt{s} = 13 \text{ TeV}$ . Tracks from the weak decay of the  $b$ -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. Baseline tracks are those produced in the standard reconstruction described in Section 3.4.1, while pseudotracks represent the ideal performance of the ATLAS detector and are described in Section 3.4.1.

These incorrect hits may skew the parameters of the track, which can in turn lower the performance of the downstream  $b$ -tagging algorithms. In particular,  $b$ -tagging algorithms rely heavily on the transverse impact parameter significance  $s(d_0)$  of the track (see Section 3.2.2). The quality of this measurement is expected to be adversely affected by wrong inner-layer hits on the track. Furthermore, multiple tracks sharing an incorrect hit can lead to the creation of spurious secondary vertices, which can cause further problems for the  $b$ -tagging algorithms.

The combination of the effects described makes reconstructing tracks in the core of high  $p_{\mathrm{T}}$   $b$ -jets particularly challenging. The reduced reconstruction efficiency of  $b$ -hadron decay tracks and incorrectly assigned hits is thought to be the primary cause of the observed drop in  $b$ -tagging efficiency at high energies, however further study is required to determine which effect may dominate.

## 4.3 High $p_{\mathrm{T}}$ $b$ -hadron Tracking Improvements

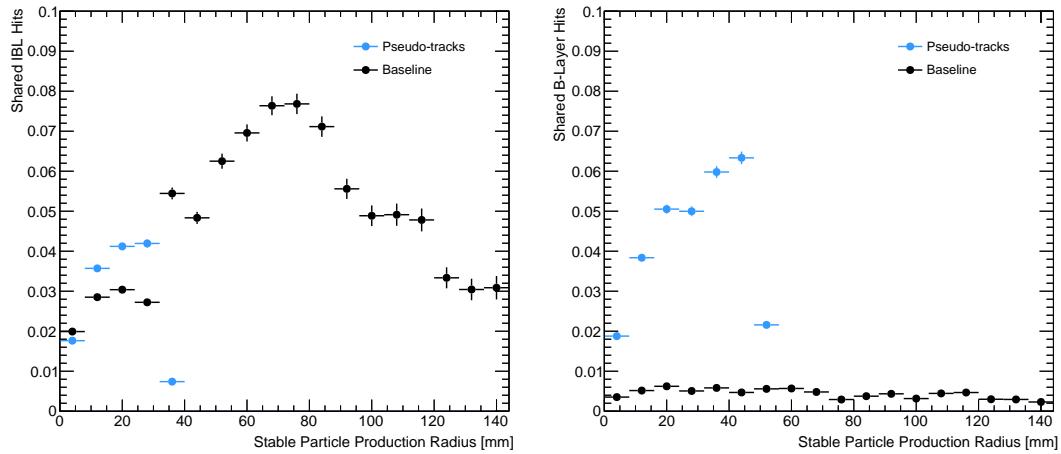
In Section 4.3.1 pseudotracks, a key tool for studying the ideal tracking performance of the ATLAS detector, are used to study the shared hit requirements on tracks in the dense cores of high- $p_{\mathrm{T}}$   $b$ -jets. Meanwhile Section 4.3.2 details a study which investigated modifying the global track fitter to improve reconstruction performance in this regime.

### 4.3.1 Shared Hits

The ambiguity solver is not run for pseudotracks. However, if the standard track collection is produced alongside the pseudotracks, then cluster splitting neural networks will be run for the standard tracks, and the resulting classification of clusters will be propagated to hits on pseudotracks. This quirk allows one to study the inefficiencies of the cluster splitting process, and relatedly to determine whether shared hit cuts in the ambiguity solver are too loose or too tight.

The fraction of hits that are shared for the IBL and the B-layer is shown in Fig. 4.4. The shared hits on pseudotracks represent correctly assigned hits from merged clusters that were not able to be classified as split by the cluster splitting neural networks.

As such, these represent the number of shared hits the ambiguity solver should aim to allow given the current performance of the cluster splitting algorithm. For shared hits on the IBL for particles produced before the IBL, the baseline selection appears to be successful in disallowing excessive numbers of shared hits. However, the ambiguity solver fails to limit shared hits for those particles produced after the IBL, reflecting the previously discussed problem of displaced tracks picking up incorrect hits. In contrast, it is clear that for the B-layer, the ambiguity solver is being overly aggressive in its rejection of shared hits, motivating further study in this area.



**Figure 4.4:** The fraction of IBL (left) and B-layer (right) hits which are shared on  $b$ -hadron decay tracks as a function of the production radius of the  $b$ -hadron decay product for tracks in jets in a  $Z'$  sample at  $\sqrt{s} = 13$  TeV. Pseudotracks represent the ideal performance given the ATLAS detector, see Section 3.4.1.

### 4.3.2 Global $\chi^2$ Fitter Outlier Removal

As part of the track fit an outlier removal procedure is run in which suspicious hits are identified and removed. This section documents ongoing studies into improving hit-to-track assignment by using the Global  $\chi^2$  Fitter (GX2F) to identify and prevent incorrect hits from being assigned to tracks during the track fit. This is in contrast to a previously investigated approach [115] which attempted to identify and remove incorrect hits after the reconstruction of the track.

The GX2F code, as a relatively low-level component of track reconstruction, has not undergone significant modification for several years, and was originally only optimised in the context of prompt, isolated tracks. Since then, a new tracking sub-detector, the IBL, was installed. The motivation for looking at the GX2F is that this change may require the re-optimisation of the GX2F code, and in particular the outlier removal procedure. Further motivation for this approach comes from the low rate of labelled outliers in baseline tracking, in contrast to the relatively higher rate of tracks with an incorrect IBL hit.

## Implementation

The outlier removal procedure for the pixel detector is described in this section. The hits on the track are looped over in order of increasing radial distance to the beam pipe. For each hit, errors  $\sigma(m_i)$  on the measurement of the transverse and longitudinal coordinates are calculated. These errors are dependent on the sub-detector which recorded the measurement (some sub-detectors are more precise than others). Additionally, a residual displacement  $r_i = m_i - x_i$  between the predicted position of the track  $x_i$  (inclusive of the current measurement), and the position of the hit itself,  $m_i$ , is calculated. The pull  $p_i$  on the track state due to the current measurement is calculated according to

$$p_i = \frac{m_i - x_i}{\sqrt{\sigma(m_i)^2 - \sigma(x_i)^2}} \quad (4.1)$$

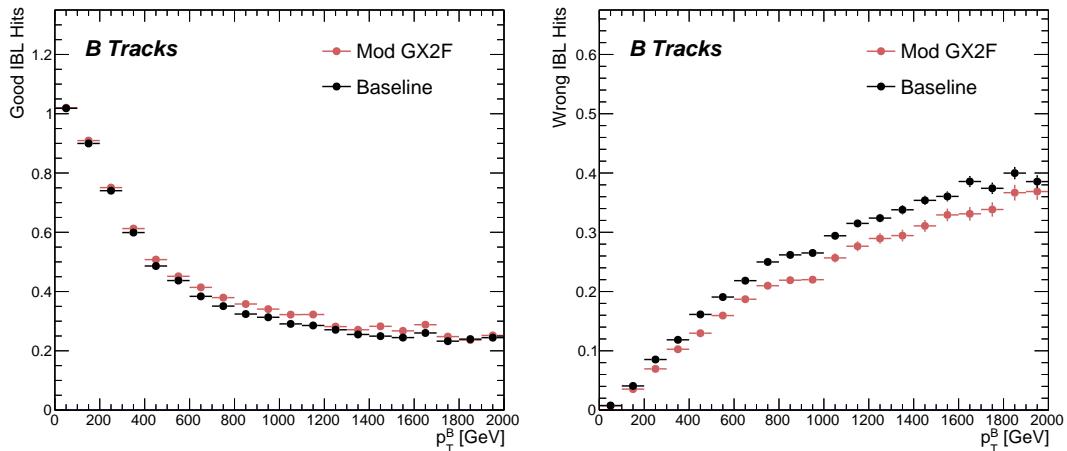
This pull is computed for the transverse and longitudinal coordinates of the measurement, and the maximum of the two is selected and checked to see if it exceeds a certain selection threshold. If it does, the hit will be removed if the track also exceeds a threshold on the total  $\chi^2/n$ , where  $n$  is the number of degrees of freedom on the track. The results of varying the outlier selection and  $\chi^2/n$  thresholds are described below.

## Selection Optimisation

A systematic variation of the outlier selection and  $\chi^2/n$  thresholds has been carried out. Both thresholds were reduced in fixed step sizes of 0.25 for the outlier selection

threshold and 1 for the  $\chi^2/n$  threshold. The value of the outlier selection threshold was reduced from 4 down to 1.75, a change which affects the silicon layers (the TRT has separate outlier removal logic). Furthermore, a specific cut for the IBL was introduced, and after optimisation is set to 1.25. The second threshold on the track  $\chi^2/n$  was also reduced from 7 to 4. Finally, instead of taking the maximum of the pulls in the longitudinal and transverse directions, a quadrature sum is taken of these two values and used. This variation is labelled “Mod GX2F” and was found to improve performance. The results for the best performing selections are discussed below.

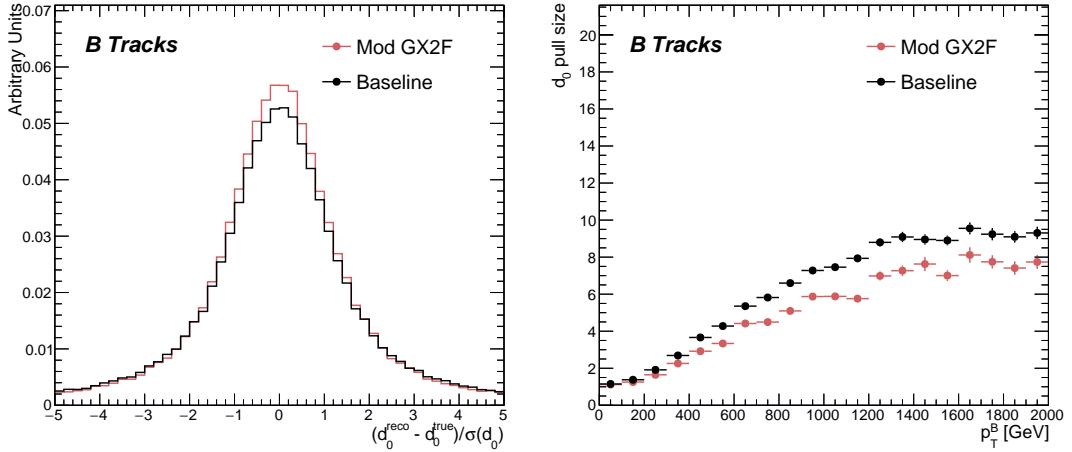
The results shown in Fig. 4.5 demonstrate a reduction in wrong hit assignment whilst also improving slightly the rate at which good hits are assigned to tracks. For a 1 TeV  $b$ -hadron, the rate to assign good hits to the corresponding decay tracks increases by approximately 10%, while the rate to assign incorrect hits decreases by approximately 16%. The improvements are also observed when looking inclusively in all tracks, which avoids the need for a specific  $b$ -jet region-of-interest selection.



**Figure 4.5:** The average number of good (left) and wrong (right) IBL hits as a function of  $b$ -hadron  $p_T$  for tracks in the  $Z'$  sample. The baseline tracking performance (black) and the modified version of the outlier removal procedure (red) are shown.

Additionally, a modest improvement of all track parameter resolutions and pulls is observed. The improvement for the transverse impact parameter pull is shown in Fig. 4.6. Note also that the large pulls for high- $p_T$   $b$ -hadrons indicates that the track parameter uncertainties are not well modelled. The results demonstrate an improvement in hit assignment, unchanged reconstruction efficiency, and modest

improvement in track parameter resolutions and pulls. In addition, the inclusive truth-matching probability of tracks is unchanged (see Section 5.2), suggesting that there is no significant increase in fake track rates. The changes are expected to have a negligible impact on computational resources.



**Figure 4.6:** (left)  $b$ -hadron decay track  $d_0$  pulls ( $d_0/s(d_0)$ ) for baseline and modified GX2F tracks in the  $Z'$  sample. (right) The absolute value of the  $d_0$  pull as a function of the truth  $b$ -hadron transverse momentum.

## 4.4 Conclusion

In this section, the difficulties facing efficient and accurate track reconstruction, and hence performant  $b$ -tagging, have been outlined. The ambiguity solver, which attempts to clean or reject tracks which have an excess of shared hits, is shown to be overly aggressive in the rejection of  $b$ -hadron decay product track candidates. The ambiguity solving process relies on a complicated pre-defined selection which has not been optimised for high transverse momentum  $b$ -hadron track reconstruction. These conclusions have motivated further ongoing studies into the improvement of the track reconstruction in dense environments and the high- $p_T$  regime, such as those in Ref. [114].

An optimisation of the outlier removal process in the global  $\chi^2$  fitter was carried out. The results of the optimisation show that more aggressive removal of outlier hits can

lead to fewer wrong hits being assigned to tracks, and improvements in the pulls of the track parameters.

## Future Work

The studies were carried out in Release 21 of the ATLAS software, and need to be reproduced using the newer Release 22 to confirm the results against other changes in the baseline tracking configuration. It is also necessary to study the impact of the improved outlier removal on the downstream  $b$ -tagging algorithms. Thanks to the all-in-one flavour tagging approach described in Chapter 6, this will in future be easier to study.

As there are some known data-MC discrepancies, fine tuned optimisation such as the work presented here presents an opportunity to over-optimise the tracking algorithms on MC. As such, further studies validating the improved outlier removal procedure on data are required.

# Chapter 5

## Track Classification NN

This chapter details work on implementing a neural network (NN) to predict the truth origin of reconstructed tracks. An introduction to the formalism of machine learning is given in Section 5.1. In Section 5.2, the truth origin label is defined, and in Section 5.3 these labels are used to train a machine learning model that can effectively discriminate between good and fake tracks. Several studies motivated this work by demonstrating that at high  $p_T$ , the performance of the low-level  $b$ -tagging algorithms was degraded by the presence of large numbers of poorly reconstructed or fake tracks. If a separate algorithm could be trained to detect fake tracks, these could be removed before their input to the  $b$ -tagging algorithms with the aim of improving performance. The identification of fake tracks could also be used to improve the reconstruction of other physics objects which rely on tracking information, for example  $\tau$  leptons or track-jets.

### 5.1 Machine Learning Background

Over the past few decades, machine learning (ML) techniques have become increasingly prevalent in High Energy Physics experiments due the increased volumes of high-dimensional data and improvements in the field of deep learning. Machine learning is the process by which a computer program uses data to infer suitable parameters for a predictive model. This is opposed to providing explicit instructions on how to perform a task. A subfield known as *supervised learning* is used in this work, and consists of exposing a model to a large number of labelled examples

in order to extract relationships between the input data and their labels. These relationships are often complex, and explicitly programmed rules can fail to fully capture the relationships between inputs and outputs.

In the simplest case, a set of  $m$  labelled training examples  $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$  is collected. Each element  $(x_i, y_i)$  consists of a input vector  $x_i \in \mathbb{R}^{\text{input}}$ , and the corresponding label  $y_i$ . In classification problems, these labels are integer *class labels*  $y_i \in \{0, \dots, N - 1\}$ , where  $N$  is the number of classes, which specify which of a pre-determined set of categorical classes the training example belongs to. The rest of the discussion in this chapter is limited to binary classification problems ( $N = 2$ ). The two classes are often referred to as signal ( $y_i = 1$ ) and background ( $y_i = 0$ ), which need to be separated. Collecting sufficient and suitable data is one of the primary challenges of machine learning, as such data is not always readily available. Fortunately, sophisticated tools to simulate particle collisions have already been developed by the scientific community [116, 117]. These tools play a key role in generating a suitably large amount of labelled data which is used to train algorithms. More detail on the input datasets can be found in Section 4.1.

After obtaining suitable training data, the next step is to define a model. Given an input domain  $\mathbb{R}^{\text{input}}$  and an output domain  $(0, 1)$ , the model  $f_\theta : \mathbb{R}^{\text{input}} \rightarrow (0, 1)$  is a parameterised functional mapping from input space to output space. Given an input example  $x_i$  and a set of parameters  $\theta$ , the model outputs a prediction  $\hat{y}_i \in (0, 1)$  for the true label  $y_i$ , as in

$$f_\theta(x_i) = \hat{y}_i. \quad (5.1)$$

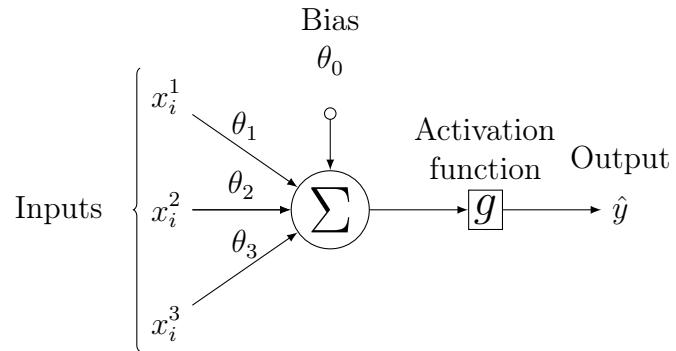
The output  $\hat{y}_i$  is in the interval  $(0, 1)$  so as to be interpreted as the probability that the input example  $x_i$  belongs to the signal class. The parameters  $\theta$  of the model are randomly initialised, and the model is designed to be expressive enough to correctly map the inputs  $x_i$  to the outputs  $y_i$  given a reasonable optimisation of the parameters. To perform this optimisation, the model is then trained, which amounts to showing the model a series of labelled training examples and modifying the parameters of the model based on its ability to correctly predict the labels.

### 5.1.1 Neural Networks

Neural networks (NNs) are a common choice for the machine learning model  $f$  since they have the ability to approximate any function [118] and are easy to train via backpropagation [119].

#### Artificial Neurons

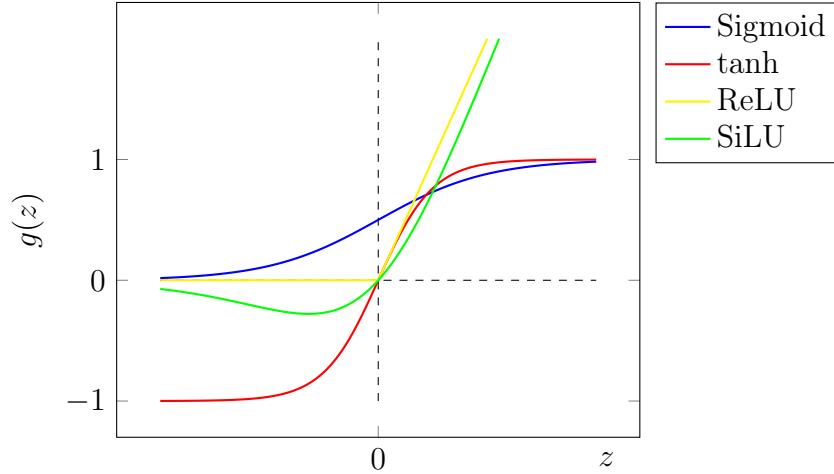
The basic functional component of a NN is the *artificial neuron* or node, which is loosely inspired by a mathematical model of a biological neuron [120, 121]. A diagram of an artificial neuron is shown in Fig. 5.1. Each neuron is defined by its parameters or *weights*  $\theta$  and a choice of activation function. Each neuron takes a fixed number of inputs and computes the dot product of the input and weight vectors  $x^T\theta$  and additionally adds a constant bias term  $\theta_0$ . This term plays the role of a trainable constant value that is independent of the inputs.



**Figure 5.1:** A diagram displaying the logical flow of a single neuron with three inputs  $x_i^j$ . Each input is multiplied by a weight  $\theta_j$ , and the resulting values are summed. A bias term  $\theta_0$  is added, and the result  $z$  is passed to an activation function. Each neuron can be thought of as a logistic regression model.

The output  $z$  of the dot product and bias term is fed into an activation function  $g(z)$ . The activation function has several uses, most notably acting as a source of non-linearity and bounding the output of the neuron. Some common activation functions (sigmoid, tanh, ReLU and SiLU) [122, 123] are shown in Fig. 5.2. The choice of activation function can have implications for the performance and convergence of the network, since the gradient of  $g(z)$  is used to compute the weight updates during

training. This is also why input data is typically normalised to have zero mean and unity variance [124].



**Figure 5.2:** The output of several common choices for the activation function  $g(z)$  of an artificial neuron. The input  $z$  is the output of the dot product between the activation and the weights, plus a bias term.

## Networks

Several neurons are linked together in layers to form a neural network. The inputs are propagated layer-by-layer through the network until reaching the final output layer. The number of layers and neurons per layer are important hyperparameters (those parameters which are not optimised as part of the training process) which influence the performance of the model. In the case of binary classification, the final output layer generally consists of a single neuron with a sigmoid activation

$$g(z) = \frac{1}{1 + e^{-z}}, \quad (5.2)$$

where  $z$  is the output from the dot product of the inputs and the weights, plus the bias term. The value  $g(z)$  is bounded between zero and one allowing the final output to be interpreted as the probability that the input sample belongs to the signal class. NNs have the crucial property of being differentiable functions, which facilitates the training process described in the next section.

### 5.1.2 Training with Gradient Descent

A training algorithm is used to optimise the weights and biases of a NN with exposure to the training data. The training algorithm works by minimising a loss function  $L$ , which quantifies the error in the model's predictions. NNs are commonly trained using backpropagation in combination with a variant of the stochastic gradient descent algorithm to iteratively update the model parameters. In binary classification problems, the binary cross entropy loss

$$L(x_i, \theta) = y_i \ln[f_\theta(x_i)] + (1 - y_i) \ln[1 - f_\theta(x_i)], \quad (5.3)$$

is often used. Since the model  $f$  is differentiable, a correction for each parameter  $\theta_i$  can be computed by taking the partial derivative of  $L$  with respect to the parameter. Updated parameters  $\theta'_i$  are calculated by updating the original parameter in the direction which reduces the loss.

$$\theta'_i = \theta_i - \alpha \frac{\partial L}{\partial \theta_i} \quad (5.4)$$

The hyperparameter  $\alpha$  is known as the *learning rate* and dictates the size of the step taken in the direction of the slope. The errors for each parameter are efficiently calculated using the backpropagation algorithm [119]. The process of updating weights is repeated until the weights are judged to have converged, which means the network is trained. In practice, small batches of the input data are shown to the network at a time. For each batch the average loss is calculated and the network's weights are updated. There are many extensions and variations of the gradient descent algorithm. This work uses the Adam optimiser which adds momentum to the weight updates (dampening oscillations) and an adaptive per-parameter learning rate [125].

## 5.2 Track Truth Origin Labelling

Crucial to supervised learning techniques are the ground truth class labels which the machine learning model is trained to predict. A set of track truth labels with a sufficient degree of granularity have been implemented in the ATLAS software stack, and are listed in Table 5.1. The labelling scheme has been designed to be

useful beyond the classification of good and fake tracks. The origins are determined by analysing the simulated record to determine the physical process that led to the creation of the truth (i.e. simulated) particle which is associated with each reconstructed track. Tracks are associated with truth particles by selecting the particle with the highest *truth-matching probability* (TMP), defined in Eq. (5.5). For a given truth particle, the TMP is a weighted sum of the number of hits on a reconstructed track which are matched to the truth particle  $N^{\text{match}}$ , divided the total number of hits on the track  $N^{\text{total}}$ . The weights are sub-detector-dependent and are designed to account for the varying importance of the different ID sub-detectors (based upon their precision) in the reconstruction of a track.

$$\text{TMP} = \frac{10N_{\text{Pix}}^{\text{match}} + 5N_{\text{SCT}}^{\text{match}} + N_{\text{TRT}}^{\text{match}}}{10N_{\text{Pix}}^{\text{total}} + 5N_{\text{SCT}}^{\text{total}} + N_{\text{TRT}}^{\text{total}}} \quad (5.5)$$

For the fake track classification tool, the track truth origins in Table 5.1 are used to construct a binary label by assigning all fake tracks to the background category, and all other tracks as signal. The fake track classifier is then trained to distinguish between these two categories of tracks. Fake tracks are defined using the TMP, with a  $\text{TMP} < 0.75$ <sup>1</sup> giving a track the label of fake. Fake tracks are made up of combinatorial fakes, which are tracks which do not correspond to the trajectory of any truth particle, and poorly reconstructed tracks, which may somewhat resemble the trajectory of a truth particle but due to the presence of some wrong hits on the track will not accurately reproduce the true trajectory. In such cases the fake track can still be identified as having an origin: it is for example possible to have a fake track which is from the decay of a  $b$ -hadron.

### 5.3 Fake Track Identification Tool

The rate of fake tracks increases at high transverse momentum as shown in Fig. 5.3 due to the difficulties in track reconstruction outlined in Section 4.2. The performance of  $b$ -tagging algorithms is reduced as a direct result of the presence of these tracks as

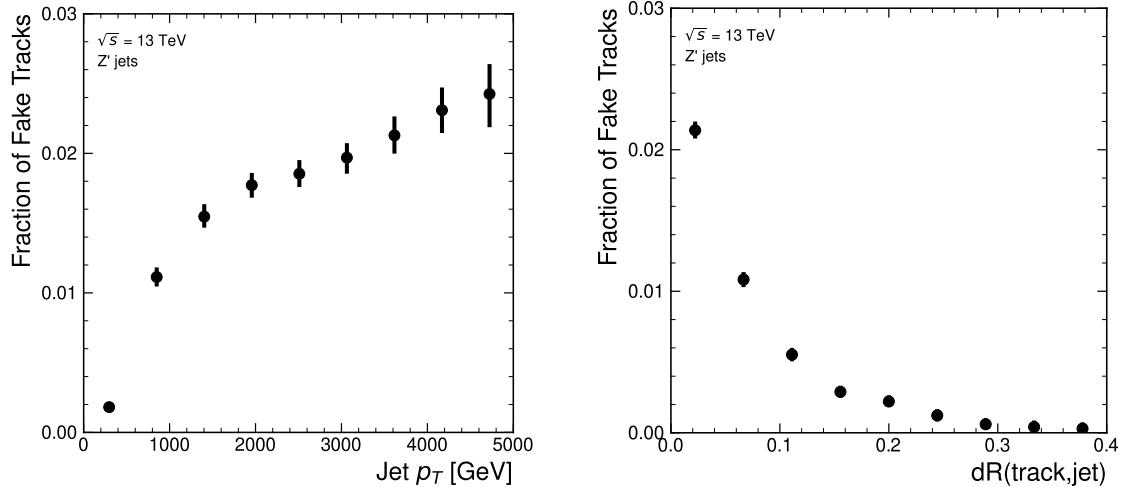
---

<sup>1</sup>An alternative definition of a fake track as one with  $\text{TMP} < 0.5$  is also in use within ATLAS, but 0.75 was used for this study.

Truth Origin	Description
Pile-up	From a $pp$ collision other than the primary interaction
Fake	Created from the hits of multiple particles
Primary	Does not originate from any secondary decay
fromB	From the decay of a $b$ -hadron
fromBC	From a $c$ -hadron decay which itself is from the decay of a $b$ -hadron
fromC	From the decay of a $c$ -hadron which is not from the decay of a $b$ -hadron
OtherSecondary	From other secondary interactions and decays

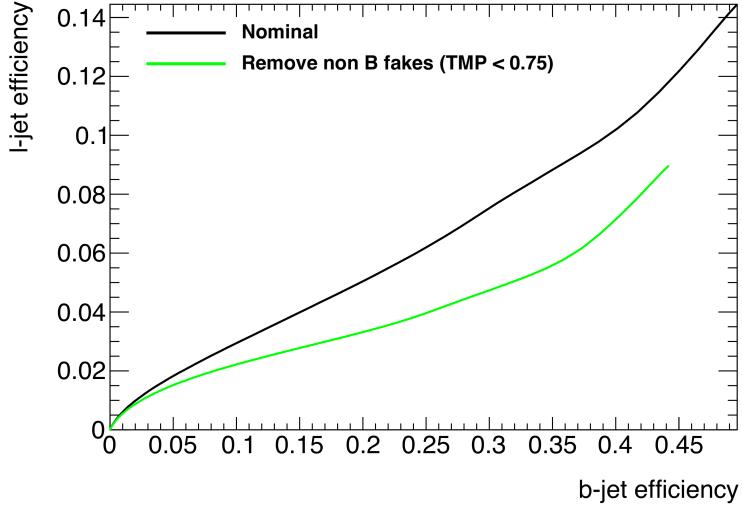
**Table 5.1:** Truth origins which are used to categorise the physical process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [63]. A truth-matching probability of less than 0.75 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The “OtherSecondary” origin includes tracks from photon conversions,  $K_S^0$  and  $\Lambda^0$  decays, and hadronic interactions.

shown for SV1 (see Section 3.4.2) in Fig. 5.4, where the efficiency to mistag a light-jet decreases by up to 35% at a  $b$ -jet efficiency of 35% if such tracks are removed.



**Figure 5.3:** Rate of fake tracks as a function of jet transverse momentum (left) and  $\Delta R(\text{track}, \text{jet})$  (right) for jets in the  $Z'$  sample. The rate of fake tracks increases significantly as a function of  $p_T$ , and also increases as the distance to the jet axis decreases.

To identify and remove fake tracks, a NN classification tool was trained with all non-fake tracks as the signal class and fake tracks as the background class. Inputs to the model are described in Section 5.3.1, while fake track removal performance is given in Section 5.3.3. Both models are trained and evaluated using tracks associated



**Figure 5.4:** The light-jet efficiency of the low level tagger SV1 for jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV, as a function of  $b$ -jet efficiency. The nominal tracking setup (black) is shown alongside the case where fake tracks which are not from the decay of a  $b$ -hadron are removed. The light-jet efficiency is decreased, demonstrating that the presence of fake tracks is detrimental to the algorithm performance.

to jets from the  $t\bar{t}$  and  $Z'$  samples described in Section 4.1. jets in the  $t\bar{t}$  sample made up 70% of the jets used for training, with the remaining 30% coming from jets in the  $Z'$  sample.

### 5.3.1 Model Inputs

The fake track NN is given two jet variables and 20 tracking related variables for each track fed into the network. The jet transverse momentum and signed pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in Table 5.2.

The track parameters and hit pattern are key indicators of whether or not a track is fake. The FracRank variable is the ordered index of the tracks that pass the ambiguity solver’s selection divided by the total number of successfully reconstructed tracks in the event. The ambiguity solver processes track candidates iteratively in order of an internal score (see Section 3.4.1), and the order in which tracks are accepted is preserved. Since tracks with shared hits have lower scores, tracks which do not require the removal of shared hits are likely to be processed and accepted

<b>Jet Input</b>	<b>Description</b>
$p_T$	Jet transverse momentum
$\eta$	Signed jet pseudorapidity
<b>Track Input</b>	<b>Description</b>
$p_T$	Track transverse momentum
$\Delta R$	Angular distance between the track and jet
$d_0$	Closest distance from the track to the PV in the longitudinal plane
$z_0$	Closest distance from the track to the PV in the transverse plane
nIBLHits	Number of IBL hits
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nTRTHits	Number of TRT hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
$r_{\text{first}}$	Radius of first hit
nDOF	Number of degrees of freedom on the track
FracRank	Ambiguity solver ordering variable

**Table 5.2:** Input features to the fake track classification NN. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [63], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split. “Primary vertex” (defined in Section 3.4.2) is abbreviated as PV.

earlier on, whereas tracks with shared hits will be processed later and potentially have their shared hits removed. Hence the FracRank variable gives an indication of the track quality and how likely it is that hits would have been removed (tracks processed later on are more likely to have hits removed).

Track selection follows the loose selection described in Ref. [97] and outlined in Table 3.2, which was found to improve the performance compared to previous tighter selections, whilst ensuring good resolution of the track’s parameters and a low fake rate [63]. Inputs are scaled to have a central value of zero and a variance of unity before training and evaluation.

### 5.3.2 Model Hyperparameters

Due to the imbalance between the two classes (with fake tracks being relatively uncommon), a weight was added to the loss function for the background class to balance their relative weights. The NN was made up of two hidden layers with 220 nodes per layer. The ReLU activation function was used in conjunction with the Adam optimiser with a learning rate of  $1\text{e}{-}3$ . Optimisation of the networks architecture was carried out to ensure optimal performance with a relatively small number of learnable parameters – 54,000. The model was trained using 40 million tracks with a further 4 million tracks each used for validation and testing. The number of tracks used for training was found to be sufficient to maximise the performance of the model, with no improvement observed when using more tracks. A full list of the model hyperparameters is given in Table 5.3.

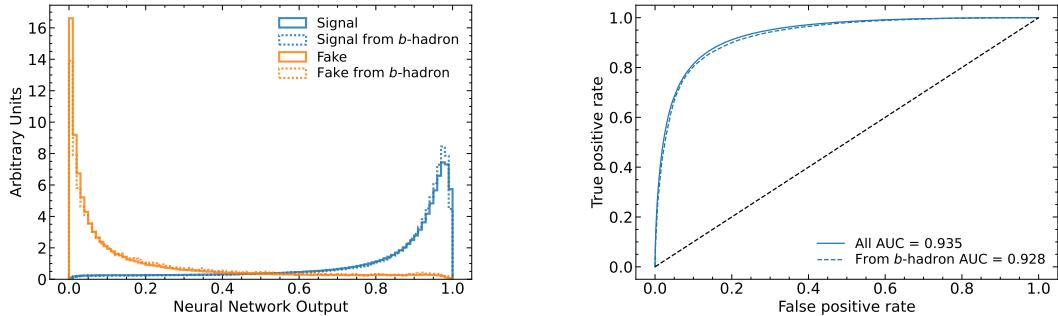
### 5.3.3 Results

In order to evaluate the fake track classification tool, a orthogonal test sample of 4 million tracks in jets in the combined  $t\bar{t}$  and  $Z'$  samples was used. The continuous scalar output from the NN model is interpreted as the probability that a given track is a signal track (i.e. not fake). Fig. 5.5 shows the performance of the fake track classification NN. The signal and background classes are well separated in the output of the tool. Also shown is a receiver operating characteristic (ROC) curve, which plots the rate of true positives against the rate of false positives over a scan of cut

Hyperparameter	Value
Batch size	2048
Activation	ReLU
Optimiser	Adam
Initial learning rate	$1e-3$
Training epochs	20
Training tracks	40m
Validation tracks	4m
Testing tracks	4m

**Table 5.3:** Hyperparameters for the track classification model.

points on the NN output ranging from zero to one. The area under the curve (AUC) gives a summary of the aggregate classification power of the model. The fake track classification tool achieves an AUC of 0.935 for all tracks, which is indicative of a well-performing model. Considering only tracks from  $b$ -hadron decays, this value drops slightly to 0.928.



**Figure 5.5:** (left) Normalised histograms of the fake track classification model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a  $b$ -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a  $b$ -hadron (dashed line). The plots show tracks in the combined  $t\bar{t}$  and  $Z'$  testing sample. The model is able to successfully discriminate between signal and fake tracks, and shows a very similar performance when looking specifically at tracks from the decay of a  $b$ -hadron.

Signal and fake track efficiencies at two different NN output cut points are shown in Table 5.4. The results demonstrate that the tool is effective in retaining 98.8% of signal tracks, while correctly identifying (and therefore enabling the removal of) 45.6% of fake tracks. Table 5.4 also shows that a significant amount of tracks which

are labelled as both fake and from the decay of a  $b$ -hadron are also removed. This can happen because fake tracks with  $\text{TMP} < 0.75$  are still matched to a truth particle, which can be the decay product of a  $b$ -hadron.

NN Output Cut	Signal Track Efficiency		Fake Track Recall	
	All	From $b$	All	From $b$
0.06	98.8%	98.9%	45.6%	39.8%
0.12	97.3%	97.5%	59.4%	53.6%

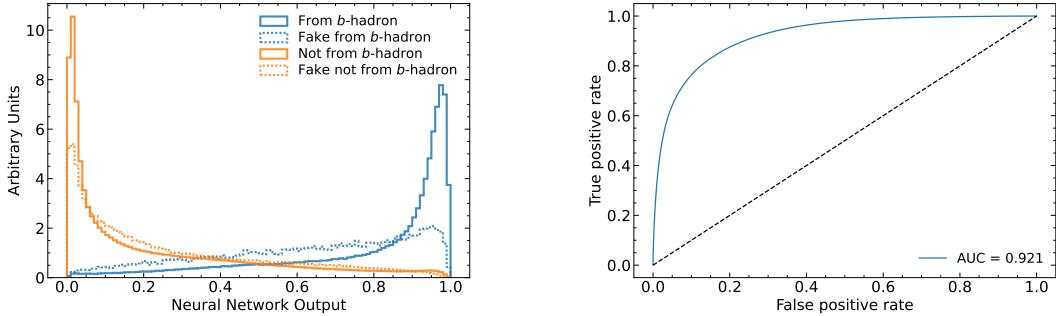
**Table 5.4:** Good and fake track selection efficiencies for the combined  $t\bar{t}$  and  $Z'$  samples. Two working points are defined, cutting on the NN output at 0.06 and 0.12.

## 5.4 $b$ -hadron Track Identification

After initial tests and investigation, it was found that fake tracks which were the result of  $b$ -hadron decays actually aided  $b$ -tagging performance, as demonstrated in Fig. 5.7. The application of a single tool which removed all fake tracks was therefore not optimal. A second tool was therefore trained in the same manner as the first, this one was designed to distinguish between those tracks which were from the decay of a  $b$ -hadron (FromB and FromBC in Table 5.1) and those which were not (all other truth origins). Fake tracks which were from the decay of a  $b$ -hadron were included in the signal class. The  $b$ -hadron decay track NN was trained using the same setup as described above, with the same tracks, input variables, and training procedure. The performance of the model to separate  $b$ -hadron decay tracks from other tracks is shown in Fig. 5.6. Using a selection WP of 0.1, the model can retain 98.5% of  $b$ -hadron tracks and reject 46.2% of tracks not from the decay of a  $b$ -hadron. In Section 5.5, this model is used in conjunction with the fake track identification NN to identify and remove fake tracks which are not from the decay of a  $b$ -hadron.

## 5.5 Combined Approach

A 2-dimensional cut was then used to only reject those tracks that had a high probability of being fake, and also a low probability of being a  $b$ -hadron decay track.



**Figure 5.6:** (left) Normalised histogram of the  $b$ -hadron track identification model output separated for tracks from the decay of a  $b$ -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line). The plots show tracks in the combined  $t\bar{t}$  and  $Z'$  testing sample.

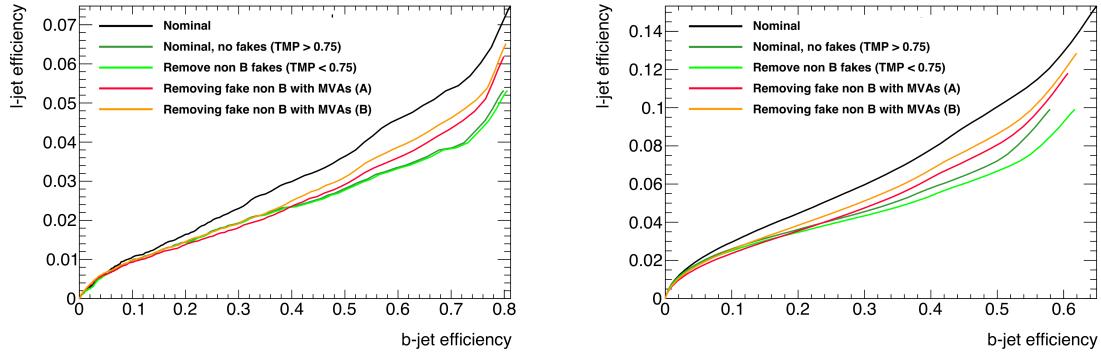
The results of the combined approach are provided in Table 5.5, which shows that for the working point “A”, 98.6% of  $b$ -hadron decay tracks (both good and fake) are retained, while 50.7% of fake tracks which are not from  $b$ -hadron decays are rejected.

WP	Fake NN Cut	$b$ -hadron Decay NN Cut	Retained $b$ -hadron Tracks	Fake non- $b$ -hadron Tracks Rejected
A	0.5	0.4	98.6%	50.7%
B	0.6	0.5	97.5%	62.0%

**Table 5.5:** Cut values for the fake and  $b$ -hadron decay track NNs for the two defined working points. Working point “B” cuts more aggressively on the NN outputs than WP “A”, removing more fake tracks but resulting in an increased loss of signal tracks (which here are all  $b$ -hadron decay tracks).

The light-jet efficiency of SV1 is successfully reduced when using the combined tools to remove fake tracks that are not from a  $b$ -hadron decay, as shown in Fig. 5.7. At a  $b$ -jet efficiency of 70%, the light-jet mistag rate for jets with  $250 < p_T < 400 \text{ GeV}$  is reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For jets with  $400 < p_T < 1000 \text{ GeV}$  the mistag rate drops from 0.1 to 0.08 for a similar relative improvement of 20%. The performance of the fake track removal approach was also tested for the other low level vertexing algorithm – JetFitter. A similar level of improvement in the light-jet mistag rate was observed with a reduction of up to a 20% reduction for both low- and high- $p_T$  jets in the  $Z'$  sample achieved. Together,

these results demonstrate that by identifying and removing fake tracks which are not the result of the weak decay of a  $b$ -hadron, the performance of the low level tagging algorithms can be improved by an amount which is comparable to the improvement that would be observed if the tracks were selected at truth level1.



**Figure 5.7:** The effect of applying the fake track identification algorithm together with the  $b$ -hadron decay track identification on the jet tagging performance of SV1 for jets in the  $Z'$  sample with  $250 \text{ GeV} < p_T < 400 \text{ GeV}$  (left) and  $400 \text{ GeV} < p_T < 1 \text{ TeV}$  (right). The nominal SV1 light-jet efficiency (black) is compared to two working points of fake track removal, labelled “A” (red) and “B” (orange), which represent two different 2D working points of the track classification tools. Removal of fake tracks based on truth information is shown by the green curves.

## 5.6 Conclusion

Fake tracks, which are prevalent in the core of high  $p_T$  jets, have an adverse impact on the performance of the low-level  $b$ -tagging algorithms SV1 and JetFitter. A ML tool to identify fake tracks has been developed. The tool can be used to limit the number of fake tracks being input to the low-level tagging algorithms. An advantage of the approach is that the continuous output of the model allows for the tuning of good and fake track identification efficiencies. Since it was found that  $b$ -hadron decay tracks can also be poorly reconstructed and thus marked as fake, it was deemed necessary also to train a second algorithm to detect  $b$ -hadron decay tracks so that the removal of these tracks could be avoided. Removing fake and non- $b$  decay tracks in this way was found to improve the light-jet mistagging rate of SV1 and JetFitter by up to 20% at high transverse momentum. The improvement achieved using the classification tools was generally comparable with that achieved when using

the truth information to remove the fake tracks not from the decay of a  $b$ -hadron, demonstrating the efficacy of the approach.

## Future Work

While removing tracks prior to their input to the low level tagging algorithms is shown here to be beneficial, a more performant alternative might be to keep these tracks but label them as being fake (for example using the output of the classification tool), and allow the tagging algorithms to take this into consideration. This is not straightforward with manually optimised low-level taggers such as SV1 and JetFitter, but is possible with more advanced taggers as described in Chapter 6.

Tools which identify the origin of a given track have other potential uses. One application is to isolate a relatively pure sample of fake tracks which can be used to estimate the fake track rate in data, which would be useful for estimating the uncertainty on fake track modelling. Another application is to use the  $b$ -hadron track identification tool to improve the track-to-jet association. Both applications are currently under investigation within ATLAS.

The approach here works on a track-by-track basis, but a more sophisticated approach would consider the correlations between the tracks inside a jet. Also left for future work is to simultaneously train a single tool which discriminates between all the truth origins listed in Table 5.1. Such a tool would be useful as a general purpose track origin classifier. An algorithm which takes both these aspects into consideration is discussed in Chapter 6.

# Chapter 6

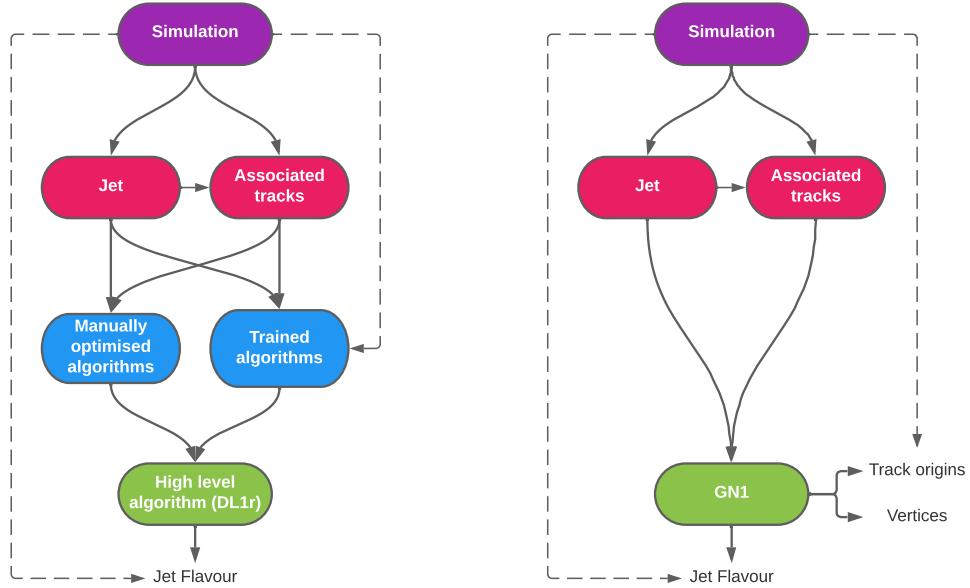
## Graph Neural Network Flavour Tagger

This chapter introduces GN1, a novel flavour tagging algorithm based on graph neural networks (GNNs). In Section 6.1, an overview of the proposed approach is given. An introduction to the formalism of GNNs is provided in Section 6.2. Details of the experimental setup are provided in Section 6.3, while a complete specification of the GN1 model is included in Section 6.4. In Section 6.5 the training procedure is described, and finally in Section 6.6 the results are shown.

### 6.1 Motivation

GN1 is a monolithic approach to flavour tagging as illustrated in Fig. 6.1. As opposed to the existing two-tiered approach to flavour tagging described in Chapter 4, which relies on the use of both low- and high-level algorithms, GN1 takes as inputs information directly from an unordered variable number of tracks, and predicts the jet flavour without requiring outputs from intermediate algorithms. In addition to predicting the flavour of the jet, the model predicts which physical processes produced the various tracks, and groups the tracks into vertices. These auxiliary training objectives provide valuable additional information about the contents of the jet and enhance the performance of the primary flavour prediction task. The GNN architecture offers a natural way to classify jets with variable numbers of

unordered associated tracks (see Section 6.2), while allowing for the inclusion of auxiliary training objectives [2, 126].



**Figure 6.1:** Comparison of the existing flavour tagging scheme (left) and GN1 (right). The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, additional truth information from the simulation is now being used as auxiliary training targets for GN1. The solid lines represent reconstructed information, whereas the dashed lines represent truth information [3].

The current flavour tagging algorithms utilise a two-tiered approach, with five low level algorithms feeding intermediate features into the high-level tagger DL1r, which outputs variables which discriminate between the different jet flavours. In contrast GN1 consists of only a single neural network, which takes the tracks as inputs along with some kinematic information about the jet. As a result, it does not depend on the outputs of any other flavour tagging algorithm. A simple training of the model fully optimises its parameters, representing a significant simplification with respect to the optimisation procedure for DL1r. This is particularly important when optimising the tagger for new regions of phase space (e.g.  $c$ -tagging or high- $p_T$   $b$ -tagging), or when the detector is upgraded or the charged particle reconstruction or selection algorithms are re-optimised.

GN1 is trained to learn about the internal structure of the jet through the use of two auxiliary training objectives: the prediction of the underlying physics process from which each track originated, and the grouping of tracks originating from a common spatial position (i.e. a common vertex). These auxiliary objectives are meant to guide the neural network towards a more complete understanding of the underlying physics inside the jet, thereby removing the need for the low-level algorithms, which previously contained information about the underlying physics in their design. The training targets for the primary and auxiliary objectives are extracted from truth information, as opposed to reconstructed quantities available in both collision data and simulation.

In this chapter, the following advantages of the GN1 approach will be demonstrated:

1. GN1 boasts improved performance with respect to the current ATLAS flavour tagging algorithms, with significantly larger background rejection rates for a given signal efficiency. Alternatively the rejection rates can be kept fixed for a substantial increase in signal efficiency, in particular at high- $p_T$ .
2. The same network architecture can be easily optimised for a wider variety of use cases (e.g.  $c$ -jet tagging and high- $p_T$  jet tagging) since there are no low-level algorithms to retune.
3. Optimal performance is achieved via the training of a single algorithm, rather than the optimisaiton of multiple low and high level algorithms.
4. Alongside the network’s prediction of the jet flavour, the auxiliary vertex and track origin predictions provide more information on why a jet was (mis)tagged or not. This information can also have uses in other applications, for instance to explicitly reconstruct displaced decay vertices or to remove fake tracks.<sup>1</sup>

## 6.2 Graph Neural Network Theory

Graph neural networks are a more sophisticated neural network model (see Section 5.1.1) that are designed to operate on graph structured data. A brief introduction to GNNs is provided in this section following the formalism in Ref. [127], while a

---

<sup>1</sup>A fake track is defined in this chapter as a track with a truth-matching probability less than 0.5, where the truth-matching probability is defined in Ref. [63].

pedagogical introduction to the usage of GNNs in particle physics can be found in Ref. [128].

A graph  $\mathcal{G}$  consists of a set of  $N^n$  nodes  $\mathcal{N} = \{h_i\}_{i=1:N^n}$ , a set of  $N^e$  edges  $\mathcal{E} = \{e_i\}_{i=1:N^e}$ , and a global representation  $u$ . Each node represents an individual object, and edges are directed connections between two nodes, called the *sender* and *receiver* nodes. The connectivity of the graph therefore encodes information about the relationships between objects that exist in the graph.

A single graph network layer consists of three separate update functions  $\phi^e$ ,  $\phi^h$  and  $\phi^u$  one for each of the nodes, edges, and global graph representation, and similarly three aggregation functions  $\rho^{e \rightarrow h}$ ,  $\rho^{e \rightarrow u}$  and  $\rho^{h \rightarrow u}$ . The aggregation functions combine information across different edges or nodes for input into the update functions, which produce new representations for the nodes, edges and global objects based on the information in the previous layer and the aggregated information. The update functions are typically each implemented as a dense feedforward neural network (as described in Section 5.1.1). The edges  $e_i$  are updated by a edge network  $\phi^e$  as in

$$e'_i = \phi^e(e_i, h_s, h_r, u), \quad (6.1)$$

where  $h_s$  and  $h_r$  are the sender and receiver nodes respectively. The nodes are updated with a node network  $\phi^h$  as in

$$h'_i = \phi^h(\bar{e}'_i, h_i, u), \quad (6.2)$$

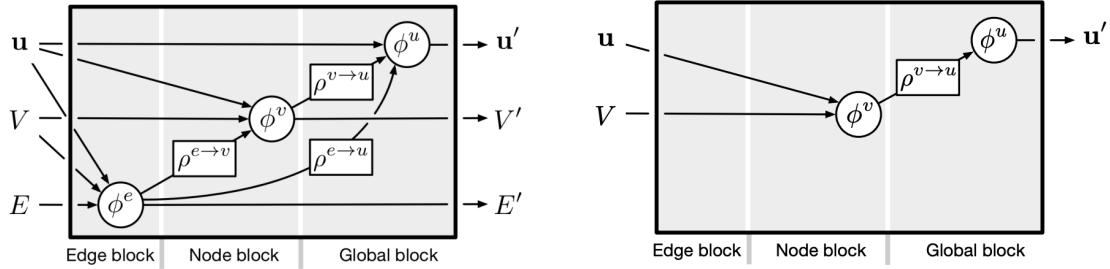
where  $\bar{e}'_i = \rho^{e \rightarrow h}(E'_i)$ , and  $E'_i$  is the set of sender nodes for receiver node  $h_i$ .  $\rho^{e \rightarrow h}$  is referred to as the edge aggregation function. The global representation is updated using the global network  $\phi^u$  as in

$$u' = \phi^u(\bar{e}', \bar{h}', u), \quad (6.3)$$

where  $\bar{e}'$  is the aggregation  $\rho^{e \rightarrow u}$  over all updated edges  $e'_i$  and  $\bar{h}'$  is the aggregation  $\rho^{h \rightarrow u}$  over all updated nodes  $h'_i$ .

The graph network layer performs a graph convolution, in an analogous way to a convolutional neural network operating on a grid of pixels. The above description is general, and not all concrete implementations of GNNs need to implement every aspect. For example, the global graph representation need not be present, and

it is also possible that no dedicated edge features are present. In such cases the corresponding update and aggregation functions are not needed. Fig. 6.2 shows two possible graph network update layers. The layer used in the GN1 model is specified in more detail in Section 6.4.3.



**Figure 6.2:** The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [129]. Reproduced from Ref. [127].

## 6.3 Experimental Setup

### 6.3.1 Datasets

The datasets used to train the GN1 tagger are the same as described in Section 4.1. The training dataset contains 30 million jets, 60% of which are  $t\bar{t}$  jets and 40% of which are  $Z'$  jets. In order to evaluate the performance of the model during training, a statistically independent validation set of 500k jets from both the  $t\bar{t}$  and  $Z'$  samples are used. For the testing of the model and the creation of the performance plots, a further 1 million independent testing jets from each of the  $t\bar{t}$  and  $Z'$  samples are used. Before being fed into the model, the track- and jet-level inputs are normalised to have a mean of zero and a variance of unity. The jet flavour labels are assigned as described in Section 3.4.3. Truth labelled  $b$ -,  $c$ - and light-jets are kinematically re-sampled in  $p_T$  and  $\eta$  to ensure identical distributions in these variables.

## 6.4 Model Specification

### 6.4.1 Model Inputs

As inputs, GN1 is fed two kinematic jet variables and an unordered set of up to 40 tracks which have been associated to the jet. If more than 40 tracks are associated to a given jet, only the first 40 tracks with the largest transverse IP significance  $s(d_0)$  (see Section 3.2.2) are fed into the model as inputs. Each track is characterised by 21 variables as detailed in Table 6.1. The kinematic jet variables are the jet transverse momentum and signed pseudorapidity. For each track, variables containing the track parameters and uncertainties, and detailed information on the hit content are provided as inputs to the model.

Dependence of the model on the absolute value of the azimuthal jet angle  $\phi$  is explicitly removed by providing only the azimuthal angle of tracks relative to the jet axis. The track pseudorapidity is also provided relative to the jet axis.

Since heavy flavour hadrons can decay semileptonically approximately 40% of the time, the presence of a reconstructed lepton in the jet carries discriminating information about the jet flavour. To exploit this, a variant of GN1 called GN1Lep is trained in addition to the baseline model. The GN1Lep variant is identical to the baseline model, except for the inclusion of an additional track-level input, leptonID, which indicates if the track was used in the reconstruction of an electron, a muon or neither. The variable is signed by the charge of the reconstructed lepton. The leptons used in the definition of the leptonID variable are required to satisfy basic quality requirements. The muons are required to be combined [130], and the electrons are required to pass the *VeryLoose* likelihood-based identification working point [131].

The selections applied to the tracks, outlined in Table 3.2, is the same as that used for the fake track classification NN described in Chapter 5. However, Section 6.7.1 demonstrates that further relaxation of the track selection requirements may be warranted.

Jet Input	Description
$p_T$	Jet transverse momentum
$\eta$	Signed jet pseudorapidity
Track Input	Description
$q/p$	Track charge divided by momentum
$d\eta$	Pseudorapidity of the track, relative to the jet $\eta$
$d\phi$	Azimuthal angle of the track, relative to the jet $\phi$
$d_0$	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin \theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on $q/p$
$\sigma(\theta)$	Uncertainty on track polar angle $\theta$
$\sigma(\phi)$	Uncertainty on track azimuthal angle $\phi$
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used to reconstruct an electron or muon

**Table 6.1:** Input features to the GN1 model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [63], while split hits are hits used on multiple tracks which have been identified as merged. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. The track leptonID is an additional input to the GN1Lep model [3]. “Primary vertex” (defined in Section 3.4.2) is abbreviated as PV.

### 6.4.2 Auxiliary Training Objectives

In addition to the jet flavour classification, two auxiliary training objectives are defined. The first auxiliary objective is the prediction of the physical process that gave rise to each track within the jet (i.e. the track origin), while the second is the prediction of track-pair vertex compatibility. Each auxiliary training objective comes with a training target which, similar to the jet flavour label, is a truth labels derived from the simulation. The presence of the auxiliary training objectives improves the jet classification performance as demonstrated in Section 6.6.3.

For the track origin prediction objective, each track is labelled with one of the exclusive categories defined in Table 5.1 of Section 5.2 after analysing the particle interaction (or lack thereof) which led to its formation. Since the presence of different track origins is strongly related to the flavour of the jet, training GN1 to recognise the origin of the tracks provides an additional handle on the classification of the jet flavour. This task may also aid the jet flavour prediction by acting as a form of supervised attention [132] - in detecting tracks from heavy flavour decays the model may learn to pay more attention to these tracks.

The vertexing auxiliary objective makes use of the fact that displaced decays of  $b$ - and  $c$ -hadrons lead to secondary and tertiary vertices inside the jet, as described in Section 3.5.2. The presence of displaced secondary vertices is not a completely clean signal of a heavy flavour jet, as displaced secondary vertices can also occur in light-jets as a result of material interactions, conversions, and long-lived particle decays (e.g.  $K_S^0$  and  $\Lambda^0$ ). For this objective, GN1 predicts a binary label for each pair of tracks in the jet. The label has a value of 1 if the truth particles associated with the two tracks in the pair originated from the same spatial point, and 0 otherwise. To derive the corresponding truth labels for training, truth production vertices within 0.1 mm are merged. Track-pairs where one or both of the tracks in the pair have an origin label of either Pile-up or Fake are assigned a label of 0. Using the pairwise predictions from the model, groups of tracks that have common compatibility can be formed, resulting in the identification of vertices. Two existing low-level tagging algorithms, SV1 and JetFitter (introduced in Section 3.4.2), are currently used to find and reconstruct vertices inside jets and are used as inputs to the existing jet flavour tagger DL1r. The addition of this auxiliary training objective removes the need for inputs from a dedicated secondary vertexing algorithm.

Both of the auxiliary training objectives described here can be considered as “stepping stones” on the way to classifying the flavour of the jet. By requiring the model to predict the truth origin of each track and the vertex compatibility of each track-pair, the model is guided to learn representations of the jet which are connected to the underlying physics and therefore relevant for classifying the jet flavour.

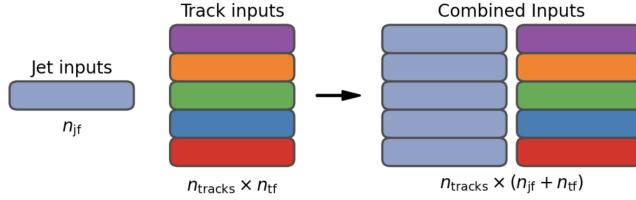
### 6.4.3 Architecture

A coarse optimisation of the network architecture hyperparameters (for example number of layers and number of neurons per layer) has been carried out in order to maximise the flavour tagging performance, but it is likely that further dedicated optimisation studies could lead to further performance improvements.

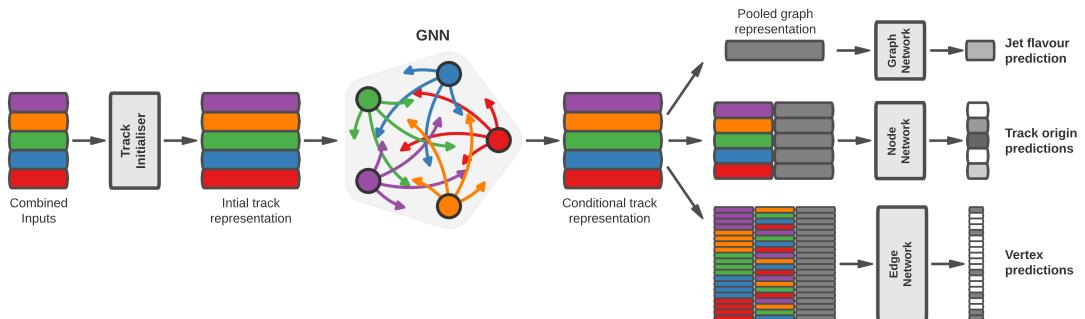
The model architecture builds on a previous implementation of a GNN-based jet tagger [126]. The previous approach was comprised of two separate graph neural networks with the auxiliary tasks being performed at an intermediate stage after the first and before the second. This two stage approach was found to be unnecessary and as such GN1 simplifies the architecture into a single graph neural network with the auxiliary tasks being performed at the end, alongside the primary jet classification task. GN1 makes use of a more sophisticated graph neural network layer [133], which is described in more detail below. The changes significantly improved tagging performance and also led to a significant reduction in training time.

As inputs, the model takes information about the jet and a number of associated tracks, as detailed in Section 6.4.1. The jet variables are concatenated with the variables for each track as shown in Fig. 6.3. The combined jet-track input vectors are then fed into a per-track initialisation network with three hidden layers, each containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 6.4. The track initialisation network is similar to a deep sets model [129], but does not include a reduction operation (mean or summation) over the output track representations. The initialisation network allows for initial per-track input processing without the associated parameter count cost of the graph convolutional layers described below.

The outputs of the track initialisation network are used to populate the nodes of a fully connected graph, such that each node in the graph neighbours every other node. Each node  $h_i$  in the graph corresponds to a single track in the jet, and is characterised



**Figure 6.3:** The inputs to GN1 are the two jet features ( $n_{jf} = 2$ ), and an array of  $n_{tracks}$ , where each track is described by 21 track features ( $n_{tf} = 21$ ). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1 [3].



**Figure 6.4:** The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting node representations are used to predict the jet flavour, the track origins, and the track-pair vertex compatibility [3].

by a feature vector, also called a representation. The per-track output representations from the initialisation networks are used as the initial feature vectors of each node in the graph. In each layer of the graph network, output node representations  $h'_i$  are computed by aggregating the features of  $h_i$  and neighbouring nodes  $\mathcal{N}_i$  using a multi-head attention mechanism ( $n = 2$ ) as described in Ref. [133, 134]. First, the feature vectors of receiver and sender nodes are fed into two fully connected linear layers  $\mathbf{W}_r$  and  $\mathbf{W}_s$ , to produce an updated representation for each sender and receiver node  $\mathbf{W}_r h_i$  and  $\mathbf{W}_s h_j$ . These updated feature vectors are used to compute edge scores  $e(h_i, h_j)$  for each node pair, as in

$$e(h_i, h_j) = \mathbf{a} \cdot \theta [\mathbf{W}_r h_i + \mathbf{W}_s h_j], \quad (6.4)$$

where,  $\theta$  is a non-linear activation function, and  $\mathbf{a}$  is a learned vector. These edge scores are then used to calculate attention weights  $a_{ij}$  for each pair of nodes using the softmax function over the edge scores

$$a_{ij} = \text{softmax}_j [e(h_i, h_j)]. \quad (6.5)$$

Finally, the updated representations for the receiver nodes  $h'_i$  are computed by taking the weighted sum over each updated node representation  $\mathbf{W}_r h_i$ , with weights  $a_{ij}$

$$h'_i = \sigma \left[ \sum_{j \in \mathcal{N}_i} a_{ij} \cdot \mathbf{W}_r h_j \right]. \quad (6.6)$$

The set of operations described above constitute a single graph network layer. Three such layers are stacked to construct the graph network, representing a balance between achieving good performance in a reasonable time and avoiding overtraining due to inflation of the parameter count of the model. The final output from the graph neural network is a set of per-node (i.e. per-track) feature vectors that are conditional representations of each track given the other tracks in the jet. In order to perform

the jet flavour prediction, a flattened global representation of the jet is needed. To produce this, the output track representations are combined using a weighted sum, where the weights are learned during training and therefore act as a form of attention over the different tracks. The flattened outputs from the sum are then fed into a fully connected feedforward neural network with four layers and three outputs, one for each jet flavour. Two other separate fully connected feedforward neural networks are then also used to independently perform the auxiliary classification objectives of GN1. A summary of the different classification networks used for the various training objectives is shown in Table 6.2.

Network	Hidden layers	Output size	Label
Node classification network	128, 64, 32	7	Track origin
Edge classification network	128, 64, 32	1	Track-pair compatibility
Graph classification network	128, 64, 32, 16	3	Jet flavour

**Table 6.2:** A summary of GN1’s different classification networks used for the various training objectives, adapted from Ref. [3]. The hidden layers column contains a list specifying the number of neurons in each layer. ReLU activation is used through the network [122].

The node classification network predicts the track truth origin as defined in Table 5.1. This network takes as inputs the features from a single output node from the graph network and the global representation of the jet. The node network has three hidden layers containing 128, 64 and 32 neurons respectively, and an output size of seven, corresponding to the seven different truth origins defined in Table 5.1.

The edge classification network is used to predict whether the tracks in the track-pair belong to a common vertex. This network takes as inputs the concatenated representations from each pair of tracks and the global jet representation. Similar to the node network, the edge network has three hidden layers containing 128, 64 and 32 neurons respectively, and a single output, which is used to perform binary classification of the track-pair compatibility. The output predictions for the two auxiliary networks are used for the auxiliary training objectives discussed in Section 6.4.2.

Finally, the graph classification network is used to predict the jet flavour. This network takes only the global jet representation as input. The graph classification

network is comprised of four fully connected hidden layers with 128, 64, 32 and 16 neurons respectively, and has three outputs corresponding to the  $b$ -,  $c$ - and light-jet classes. To obtain probability outputs for each task, the outputs from each classification network are passed through a softmax function.

## 6.5 Training

The full GN1 training procedure minimises the total loss function  $L_{\text{total}}$ , defined as

$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}}. \quad (6.7)$$

This loss is composed of three terms:  $L_{\text{jet}}$ , the categorical cross entropy loss over the different jet flavours;  $L_{\text{vertex}}$ , the binary track-pair compatibility cross entropy loss; and  $L_{\text{track}}$ , the categorical cross entropy loss for the track origin prediction.  $L_{\text{vertex}}$  is computed via a weighted average over all intra-jet track-pairs in the batch, and  $L_{\text{track}}$  is computed by a weighted average over all tracks in the batch, where the weights are described below.

The different losses converge to different values during training, reflecting differences in the relative difficulty of the various training objectives. The values of  $L_{\text{vertex}}$  and  $L_{\text{track}}$  are weighted by  $\alpha = 1.5$  and  $\beta = 0.5$  respectively to ensure they converge to similar values, giving them an equal weighting towards  $L_{\text{total}}$ . The values of  $\alpha$  and  $\beta$  are chosen to ensure that  $L_{\text{jet}}$  converges to a larger value than either  $L_{\text{vertex}}$  and  $L_{\text{track}}$ , which reflects the primary importance of the jet classification objective. It was found that in practice the overall performance of the model was not sensitive to modest changes in the loss weights  $\alpha$  and  $\beta$ . Pre-training using  $L_{\text{total}}$  (i.e. on all tasks) and fine tuning on only the jet classification task also did not improve performance versus the standard setup, indicating that the auxiliary tasks are not in direct competition with the jet classification task. As there was a large variation in the relative abundance of tracks of the different origins, the contribution of each origin to  $L_{\text{track}}$  was weighted by the inverse of the frequency of their occurrence. In vertexing loss  $L_{\text{vertex}}$ , the class weight for track-pairs where both tracks are from either a  $b$ - or  $c$ -hadron was increased by a factor of two as compared with other

track-pairs, to encourage the network to focus on correctly classifying heavy flavour vertices.

GN1 can be trained with either the node or edge networks (and their corresponding auxiliary tasks), or both, removed, as discussed in Section 6.6.3. In such cases, the corresponding losses  $L_{\text{vertex}}$  and  $L_{\text{track}}$  are also removed from the calculation of the overall loss  $L_{\text{total}}$ . The performance of the resulting models provides a useful indication of the benefit of including the auxiliary tasks to the primary jet classification objective.

GN1 was trained for 100 epochs on 4 NVIDIA V100 GPUs. Each epoch takes approximately 25 mins to complete over the training sample of 30 million jets. The Adam optimiser [135] with an initial learning rate of  $1\text{e}{-}3$ , and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically the validation loss, calculated on 500k jets, became stable after around 60 epochs. The epoch that minimized the validation loss was used for evaluation. GN1 has been integrated into the ATLAS software [59] using ONNX [136]. The test sample jet flavour predictions scores are computed using the ATLAS software stack as a verification of this process.

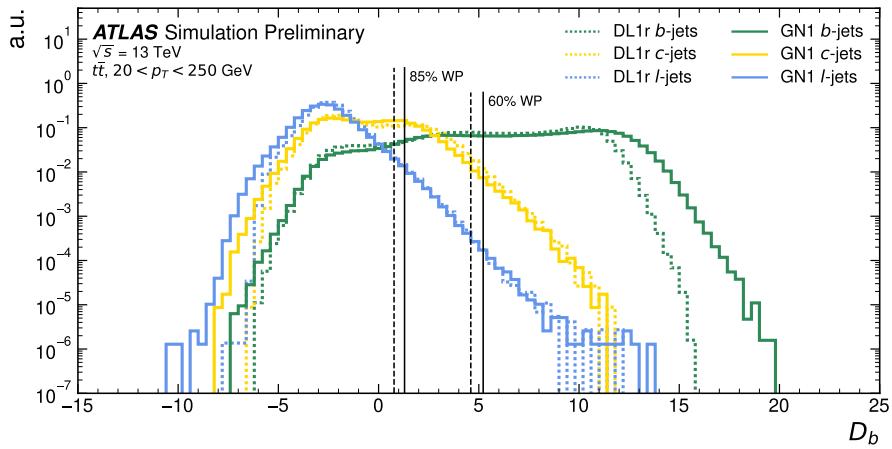
## 6.6 Results

The GN1 tagger is evaluated both as a  $b$ -tagging and  $c$ -tagging algorithm in Section 6.6.1 and Section 6.6.2 respectively. Evaluation is performed separately on jets in the  $t\bar{t}$  sample with  $20 < p_{\text{T}} < 250 \text{ GeV}$  and jets in the  $Z'$  sample with  $250 < p_{\text{T}} < 5000 \text{ GeV}$ . The performance of the model is compared to the DL1r tagger [99, 100], which has been retrained on 75 million jets from the same samples as GN1. The input RNNIP tagger [96] to DL1r has not been retrained. As discussed, each tagger predicts the probability that a jet belongs to the  $b$ -,  $c$ - and light-classes. To use the model for  $b$ -tagging, these probabilities are combined into a single score  $D_b$ , which is defined as

$$D_b = \log \frac{p_b}{(1 - f_c)p_l + f_c p_c}, \quad (6.8)$$

where  $f_c$  is a free parameter that determines the relative weight of  $p_c$  to  $p_l$  in the score  $D_b$ , controlling the trade-off between  $c$ - and light-jet rejection performance. The choice of  $f_c$  is arbitrary, and is optimised based upon the desired light- vs  $c$ -jet rejection performance. This parameter is set to a value of  $f_c = 0.018$  for the DL1r model, obtained through an optimisation procedure described in Ref. [99]. Based on a similar optimisation procedure, a value of  $f_c = 0.05$  is used for the GN1 models.

A comparison of the  $b$ -tagging discriminant  $D_b$  between DL1r and GN1 is shown in Fig. 6.5. The shapes of the  $D_b$  distributions are generally similar for  $b$ -,  $c$ - and light-jets between both models, however, GN1 shifts the  $b$ -jet distribution to higher values of  $D_b$  in the regions with the greatest discrimination. The GN1  $c$ -jet distribution is also shifted to lower values of  $D_b$  when compared with DL1r, enhancing the separation and indicating that GN1 is improving  $c$ -jet rejection when compared with DL1r.



**Figure 6.5:** Comparison between the DL1r and GN1  $b$ -tagging discriminant  $D_b$  for jets in the  $t\bar{t}$  sample. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1. The distributions of the different jet flavours have been normalised to unity area [3].

### 6.6.1 $b$ -tagging Performance

The performance of  $b$ -tagging algorithms is quantified by their ability to reject  $c$ - and light-jets for a given  $b$ -jet selection efficiency WP. In order to compare the  $b$ -tagging performance of the different taggers for the  $b$ -jet tagging efficiencies in the range

typically used by analyses, the corresponding  $c$ - and light-jet rejection rates are displayed in Figs. 6.6 and 6.7 for jets in the  $t\bar{t}$  and  $Z'$  samples respectively. Four standard WPs are defined with  $b$ -jet tagging efficiencies of 60%, 70%, 77% and 85% respectively. These WPs are commonly used by physics analyses depending on their specific signal and background requirements. The WPs are defined based on jets in the  $t\bar{t}$  sample only. Due to the much higher jet  $p_T$  range in the  $Z'$  sample, and the increased difficulty in tagging jets at high- $p_T$  (see Chapter 4), the corresponding  $b$ -jet tagging efficiencies for jets in the  $Z'$  sample are lower than the corresponding WPs calculated in the  $t\bar{t}$  sample. For instance the WP cut value computed to provide a 70%  $b$ -jet tagging efficiency on the  $t\bar{t}$  sample results in a  $b$ -jet tagging efficiency of just  $\sim 30\%$  on the  $Z'$  sample. In order to account for this, the range of  $b$ -jet tagging efficiencies displayed for plots showing the performance for jets in the  $Z'$  sample (for example Fig. 6.7) is chosen to span the lower efficiencies achieved in the  $Z'$  sample at high- $p_T$ .

For jets in the  $t\bar{t}$  sample with  $20 < p_T < 250 \text{ GeV}$ , GN1 demonstrates considerably better  $c$ - and light-jet rejection when compared with DL1r across the full range of  $b$ -jet tagging efficiencies studied. The relative improvement is strongly dependent on the  $b$ -jet tagging efficiency under study. The largest improvements are found at lower  $b$ -jet tagging efficiencies. At a  $b$ -jet tagging efficiency of 70%, the  $c$ -jet rejection improves by a factor of  $\sim 2.1$  while the light-jet rejection improves by a factor of  $\sim 1.8$  with respect to DL1r. For high- $p_T$  jets in the  $Z'$  sample with  $250 < p_T < 5000 \text{ GeV}$ , GN1 also brings a significant performance improvement with respect to DL1r across the range of  $b$ -jet tagging efficiencies studied. Again, the largest relative improvement in performance comes at the lower  $b$ -jet tagging efficiencies. At a  $b$ -jet efficiency of 30%, GN1 improves the  $c$ -jet rejection with respect to DL1r by a factor of  $\sim 2.8$  and the light-jet rejection by a factor of  $\sim 6$ . The performance comparison at lower  $b$ -jet tagging efficiencies is made more difficult due to the increased statistical uncertainties which result from the high rejection of background.

The GN1Lep variant of GN1 demonstrates further improved performance with respect to the baseline model. This demonstrates the additional jet flavour discrimination power provided by the leptonID track input. For jets in the  $t\bar{t}$  sample, the relative  $c$ -jet rejection improvement with respect to GN1 at the 70%  $b$ -jet WP is approximately 25%. The improvement in light-jet rejection also increases by 40% at the same WP. For jets in the  $Z'$  sample, the relative  $c$ -jet rejection (light-jet rejection) performance

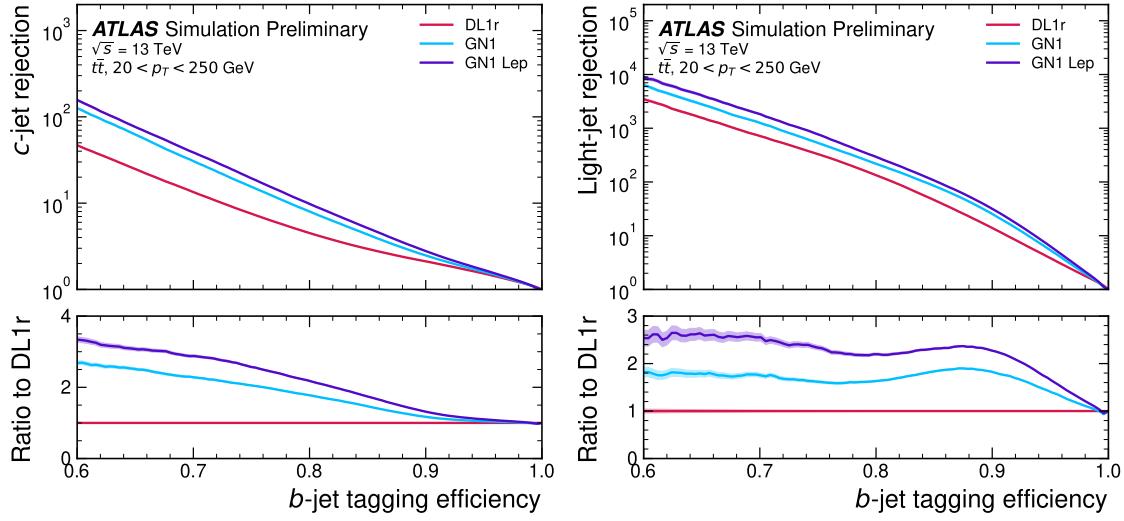
with respect to GN1 improves by approximately 10% (25%) at a  $b$ -jet tagging efficiency of 30%.

In general, the performance of all the taggers is strongly dependent on jet  $p_T$ , due to the increased multiplicity and collimation of tracks, and the displaced decays that result from within the heavy flavour jets (see Chapter 4). Together, they contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a general degradation in quality of tracks inside the core of a jet, which in turn reduces the jet tagging performance. In order to study how the tagging performance changes as a function of the jet  $p_T$ , the  $b$ -jet tagging efficiency as a function of  $p_T$  for a fixed light-jet rejection of 100 in each bin is shown in Fig. 6.8. For jets in the  $t\bar{t}$  sample, at a fixed light-jet rejection of 100, GN1 improves the  $b$ -jet tagging efficiency by approximately 4% across all the jet  $p_T$  bins. Meanwhile, GN1Lep again demonstrates improved performance with respect to GN1, in particular at lower  $p_T$ . The relative increase in the  $b$ -jet tagging efficiency increases from 4% to 8% with respect to DL1r. For jets in the  $Z'$  sample, GN1 again outperforms DL1r across the entire jet  $p_T$  range studied. The largest relative improvement in performance is found at the highest transverse momenta of jet  $p_T > 2 \text{ TeV}$ , and corresponds to an approximate factor of 2 improvement in efficiency with respect to DL1r.

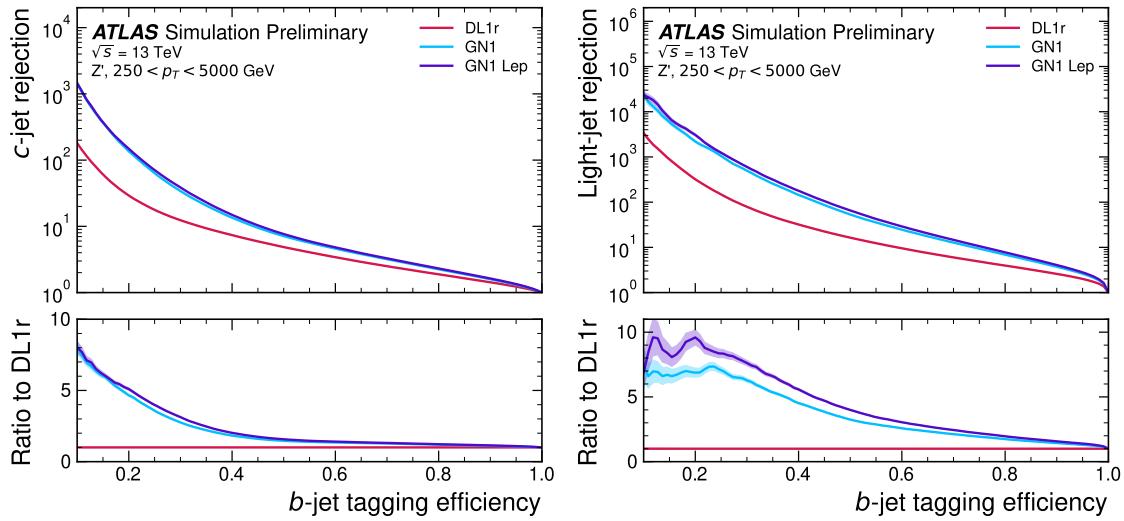
The performance of the model was also evaluated as a function of the average number of pile-up interactions in the event. No significant dependence of the tagging performance was observed.

### 6.6.2 $c$ -tagging Performance

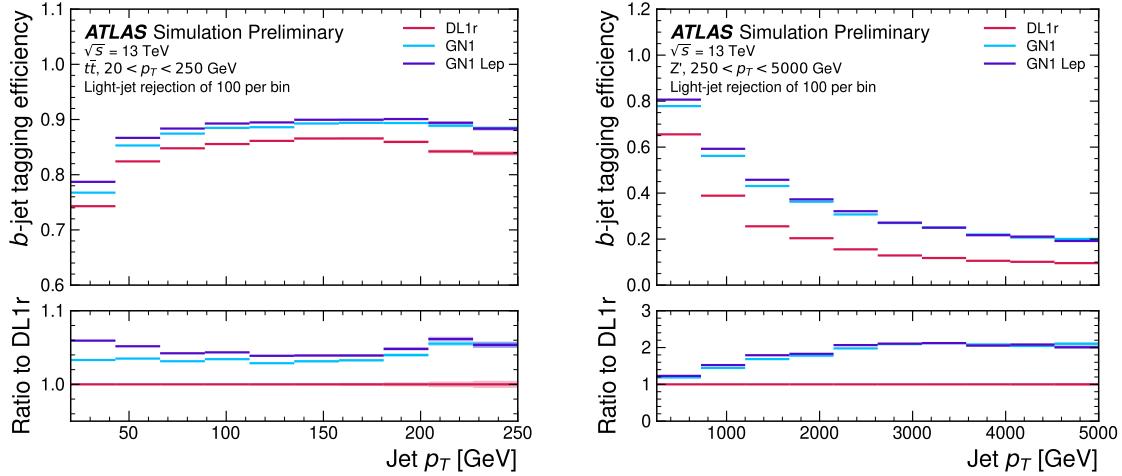
As discussed previously, GN1 does not rely on any inputs from manually optimised low-level tagging algorithms. Since these algorithms were originally designed and tuned with the aim of  $b$ -tagging, and not  $c$ -tagging, the low level tagging algorithms may perform suboptimally for  $c$ -tagging purposes. The tagging of  $c$ -jets therefore presents a compelling use case for GN1. As each of the models is trained with three output classes, using it as a  $c$ -tagging algorithm is trivially analogous to the approach used for  $b$ -tagging. The model output probabilities are combined into a single score  $D_c$ , which is defined similarly to Eq. (6.8) as



**Figure 6.6:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250 \text{ GeV}$ . The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $b$ -jet tagging efficiencies usually probed in these regions of phase space [3].



**Figure 6.7:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for jets in the  $Z'$  sample with  $250 < p_T < 5000 \text{ GeV}$ . The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $b$ -jet tagging efficiencies usually probed in these regions of phase space [3].



**Figure 6.8:** The  $b$ -jet tagging efficiency for jets in the  $t\bar{t}$  sample (left) and jets in the  $Z'$  sample (right) as a function of jet  $p_T$  with a fixed light-jet rejection of 100 in each bin. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. GN1 demonstrates improved performance with respect to DL1r across the  $p_T$  range shown. Binomial error bands are denoted by the shaded regions [3].

$$D_c = \log \frac{p_c}{(1 - f_b)p_l + f_b p_b}. \quad (6.9)$$

A value of  $f_b = 0.2$  is used for all models, based on the same optimisation procedure that was used for the  $b$ -tagging use case. Similar to Section 6.6.1, the different taggers are compared to one another by scanning through a range of  $c$ -jet tagging efficiencies and plotting the corresponding  $b$ - and light-jet rejection rates, and the WPs are defined using jets in the  $t\bar{t}$  sample. Standard  $c$ -jet tagging efficiency WPs used by physics analyses are significantly lower than the  $b$ -tagging WPs in order to maintain reasonable  $b$ - and light-jet rejection rates. This is reflected in the range of  $c$ -jet tagging efficiencies used in  $c$ -tagging plots such as Figs. 6.9 and 6.10. Fig. 6.9 displays the  $c$ -tagging performance of the models on the jets in the  $t\bar{t}$  sample. GN1 is shown to perform significantly better than DL1r. Similar to the  $b$ -tagging case, the  $b$ - and light-jet rejection improve most at lower  $c$ -jet tagging efficiencies, with the  $c$ -jet rejection (light-jet rejection) improving by a factor 2 (1.6) with respect to DL1r at a  $c$ -jet tagging efficiency of 25%. GN1Lep again outperforms GN1, though the improvements are more modest than observed for the  $b$ -tagging use case,

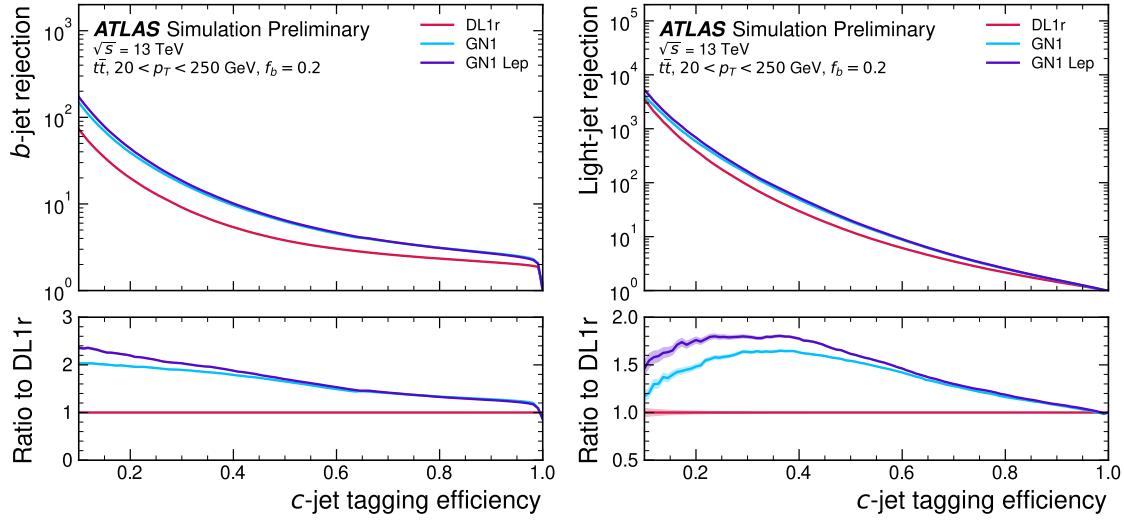
with both the  $b$ -jet rejection (light-jet rejection) improving with respect to GN1 by approximately 10% (20%) at the 25%  $c$ -jet WP. Fig. 6.10 shows the  $c$ -tagging performance on the jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV. Both GN1 and GN1Lep perform similarly, improving the  $b$ -jet rejection by 60% and the light-jet rejection by a factor of 2 at the 25%  $c$ -jet WP.

### 6.6.3 Ablations

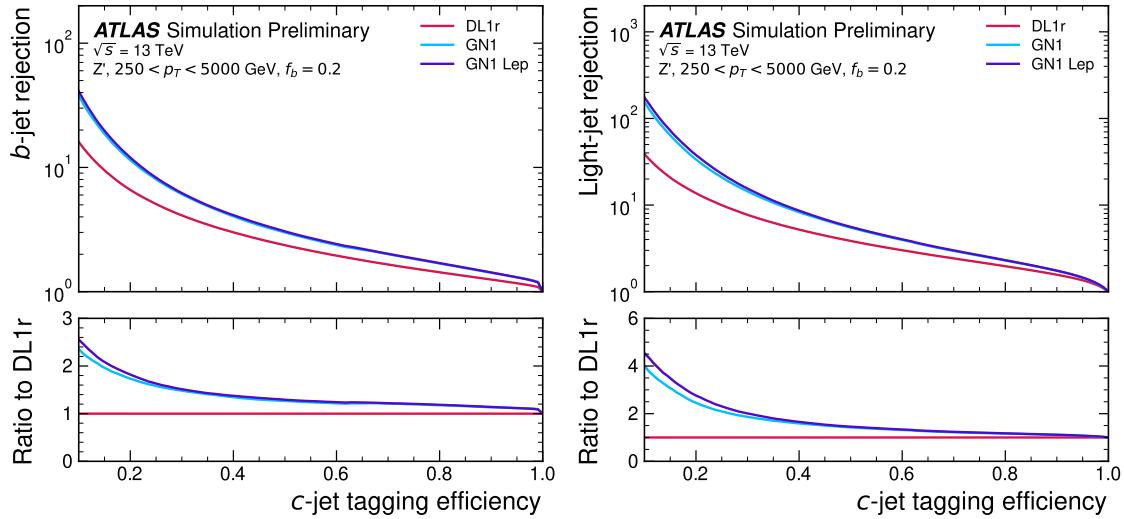
Ablation studies (the removal of certain components of a model in order to study the impact of that component) are carried out to determine the importance of the auxiliary training objectives of GN1 to the overall performance. The “GN1 No Aux” variant retains the primary jet classification objective, but removes both track classification and vertexing auxiliary objectives and correspondingly only minimises the jet classification loss. The “GN1 TC” variant includes the track classification objective but not the vertexing objective. Finally, the “GN1 Vert” includes the vertexing objective, but not the track classification objective.

For jets in both the  $t\bar{t}$  and  $Z'$  samples, a general trend is observed that the models trained without one or both of the auxiliary objectives results in significantly reduced  $c$ - and light-jet rejection when compared with the baseline GN1 model, as shown in Figs. 6.11 and 6.12. For jets in the  $t\bar{t}$  sample, the performance of GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert perform similarly to each other. For jets in the  $Z'$  sample, the GN1 No Aux model shows a clear improvement in  $c$ - and light-jet rejection when compared with DL1r at lower  $b$ -jet tagging efficiencies. Similar to jets in the  $t\bar{t}$  sample, GN1 TC and GN1 Vert perform similarly, and bring large gains in background rejection when compared with GN1 No Aux, but the combination of both auxiliary objectives yields the best performance.

It is notable that the GN1 No Aux model matches or exceeds the performance of DL1r without the need for inputs from the low-level algorithms. This indicates that the performance improvements enabled by the improved neural network architecture used in GN1 appear to be able to compensate for the removal of the low-level algorithm inputs. The GN1 TC and GN1 Vert variants each similarly outperform DL1r, demonstrating that both contribute to the overall high performance of the baseline model. The overall best performing model is the full version of GN1 trained



**Figure 6.9:** The  $b$ -jet (left) and light-jet (right) rejections as a function of the  $c$ -jet tagging efficiency for  $t\bar{t}$  jets with  $20 < p_T < 250 \text{ GeV}$ . The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $c$ -jet tagging efficiencies usually probed in these regions of phase space [3].



**Figure 6.10:** The  $b$ -jet (left) and light-jet (right) rejections as a function of the  $c$ -jet tagging efficiency for  $Z'$  jets with  $250 < p_T < 5000 \text{ GeV}$ . The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $c$ -jet tagging efficiencies usually probed in these regions of phase space [3].

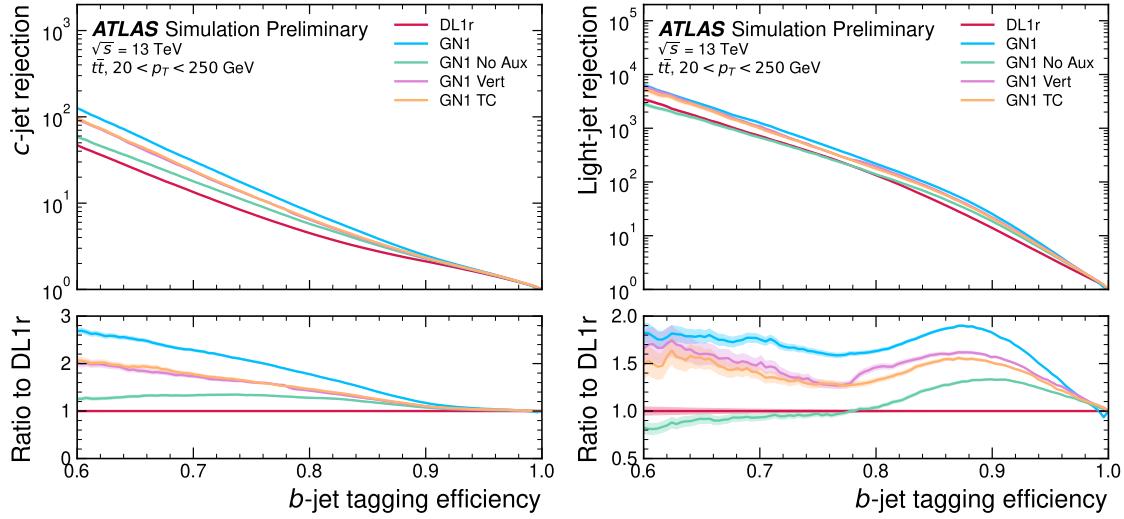
with both auxiliary objective, demonstrating that the two auxiliary objectives are complementary.

#### 6.6.4 Inclusion of Low-Level Vertexing Algorithms

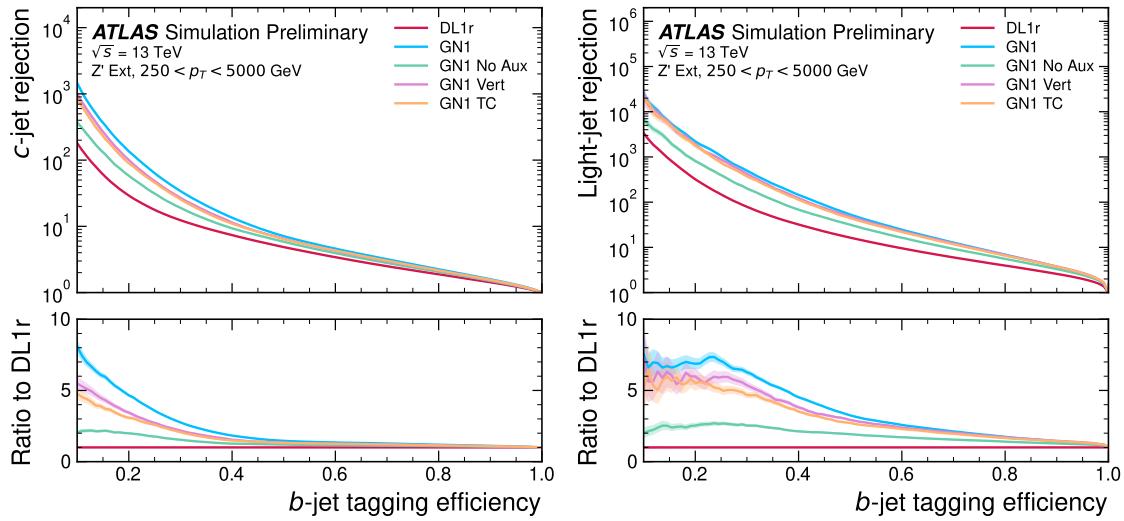
As already mentioned, GN1 does not include any inputs from the low-level tagging algorithms, including the vertexing algorithms SV1 and JetFitter [68]. Since these algorithms are known to play a key role in contributing to the performance of DL1r, it was studied whether their inclusion in GN1 could result in further performance improvements. In a dedicated training of GN1, the SV1 and JetFitter tagger outputs were added to the GN1 jet classification network as an input, similar to how they are used in DL1r. These outputs include information on the reconstructed vertices, including the number of vertices, and the properties of the reconstructed vertices. In addition, if track was used in the reconstruction of a vertex, a corresponding index to the vertex was included as a track-level inputs to GN1. These indices were also used to construct an input feature for the edge classification network used to identify vertices, which was given a value of one if the track-pair were from a common reconstructed SV1 or JetFitter vertex, and zero otherwise. The jet classification performance of this GN1 model was not significantly different to the baseline model, and in some cases the performance was slightly reduced. It was therefore concluded that GN1 does not benefit from the inclusion of information from SV1 and JetFitter, indicating that the model is able to reconstruct the relevant information provided by these low-level algorithms. The study also demonstrates that the model can function as a highly performant standalone tagger that does not require (beyond retraining) any manual optimisation to achieve good performance in a wide range of phase spaces. A dedicated look at the vertexing performance of GN1 with some comparisons to SV1 and JetFitter is found in Section 6.6.6

#### 6.6.5 Jet Display Diagrams

The auxiliary training objectives of GN1 allow for improved model interpretability, which is especially important for a monolithic approach as the low level taggers, which provide useful physical insight, are no longer present. Figs. 6.13 and 6.14 provide example comparisons of the true origin and vertexing information compared



**Figure 6.11:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $t\bar{t}$  jets with  $20 < p_T < 250$  GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 [3].



**Figure 6.12:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $Z'$  jets with  $250 < p_T < 5000$  GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 [3].

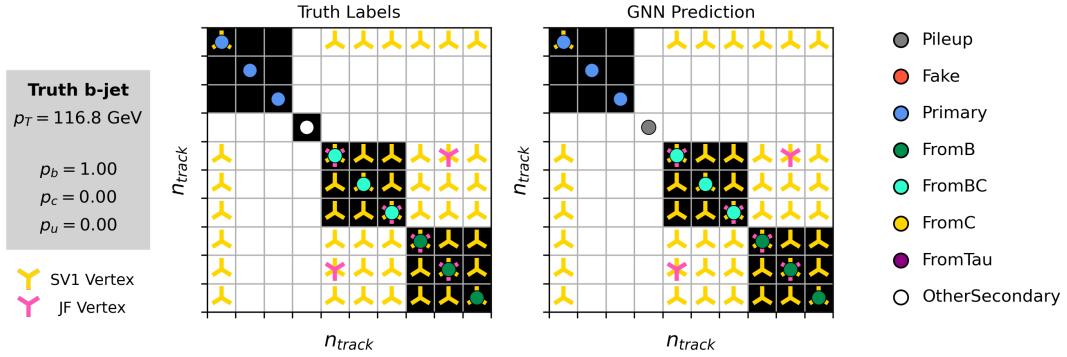
with the predicted values from GN1, SV1 and JetFitter. Such comparisons can be used to provide an indication that GN1 reconstructs the correct representation of the jet structure, and may also help to identify limitations of the model. In the figures, the tracks in the jet are indexed twice on each of the  $x$ - and  $y$ -axes, and tracks are grouped into vertices along with other tracks as indicated by common markings in the relevant rows and columns.

In Fig. 6.13, GN1 correctly groups the three primary tracks as having come from the primary vertex. The  $b$ -hadron and  $b \rightarrow c$ -hadron decay vertices are also correctly predicted, and the origin of the tracks in each is correct. There is a single OtherSecondary track which GN1 incorrectly predicts as having come from pile-up. Meanwhile SV1 (by design) merges the two heavy flavour decay vertices, but incorrectly includes a track from the primary vertex. JetFitter reconstructs two vertices, one which is a combination of two tracks from different truth vertices and two other single track vertices in each of the heavy flavour vertices. GN1 also predicts the flavour of the jet with a high degree of certainty.

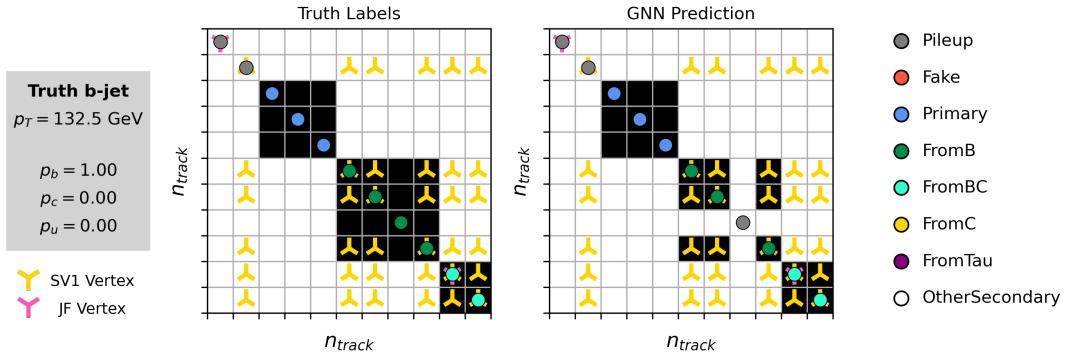
Similarly Fig. 6.13 shows that GN1 is also able to relatively accurately predict the origin and vertex information of tracks inside a jet. The pile-up tracks and primary vertex tracks are correctly identified, and the heavy flavour decay tracks are also correctly identified with the exception of one of the  $b$ -hadron decay tracks. Again, SV1 merges the two heavy flavour decay vertices along with a track from pile-up, while JetFitter shows signs of being under-constrained by reconstructing two single track vertices, one with a pile-up track and one with a track from a  $b \rightarrow c$ -hadron decay. While these examples do not give a complete picture of the performance of GN1, they do show provide a powerful way to visualise and diagnose the behaviour of GN1.

### 6.6.6 Vertexing Performance

From the track-pair vertex prediction, tracks can be partitioned into compatible groups representing vertices through the use of a deterministic clustering algorithm (see Ref. [126]). As such, GN1 can perform vertex “finding”, but not vertex “fitting”, i.e. the reconstruction of a vertex’s properties, which currently still requires the use of a dedicated vertex fitter. In order to study the performance of the different vertexing tools, the truth vertex label of the tracks are used.



**Figure 6.13:** A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a  $b$ -jet. The true and predicted origins of the tracks is shown by the coloured circles along the diagonal. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities  $p_b$ ,  $p_c$  and  $p_u$  are rounded to three significant figures. GN1 improves over SV1 and JetFitter by successfully determining the origins and vertex compatibility of all the tracks in the jet with the exception of the truth OtherSecondary track, which is misidentified as pile-up.



**Figure 6.14:** A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a  $b$ -jet. The true and predicted origins of the tracks is shown by the coloured circles along the diagonal. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities  $p_b$ ,  $p_c$  and  $p_u$  are rounded to three significant figures. GN1 improves over SV1 and JetFitter by successfully determining the origins and vertex compatibility of all but one tracks in the jet.

There are several caveats to a comparison of the vertexing tools which are a result of the different approaches they take to vertexing. SV1 and JetFitter are designed to only find secondary vertices in the jet, whereas GN1 is also trained to determine which tracks in the jet belong to the primary vertex. To account for this the GN1 vertex with the largest number of predicted primary tracks is excluded from the vertex finding efficiency calculation. While JetFitter and GN1 aim to resolve each displaced vertex inside the jet, such that secondary vertices from  $b$ -hadron decays are found separately to tertiary vertices from  $b \rightarrow c$  decay chains, SV1 by design attempts to find a single inclusive vertex per jet. This inclusive vertex groups tracks from the  $b$ -hadron decay itself (FromB) and tracks from  $b \rightarrow c$  decays (FromBC). In order to fairly compare the performance of the different tools, both the exclusive and inclusive vertex finding efficiencies are studied. For the exclusive vertex finding case JetFitter and GN1 can be directly compared, while a comparison with SV1 is not possible due to the aforementioned design constraints. The inclusive vertex finding performance of all three tools can be compared using the procedure outlined below.

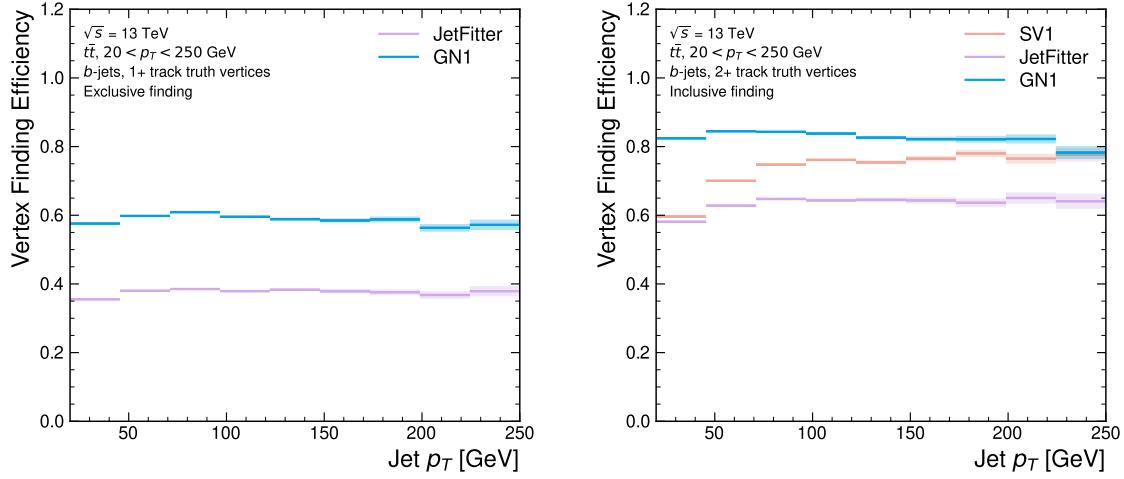
The starting point for the secondary vertex finding efficiency in both the exclusive and inclusive cases is to select truth heavy flavour (HF) secondary vertices, defined as those containing only inclusive  $b$ -hadron decays. For exclusive HF vertex finding, these truth secondary HF vertices can be used directly as the denominator for the efficiency calculation. Meanwhile for the inclusive efficiency all such truth HF secondary vertices in the jet are merged into a single inclusive target vertex. Correspondingly, for the inclusive HF vertex finding case, the vertices found by JetFitter are merged into a single vertex, and the vertices found by GN1 which contain at least one predicted  $b$ -hadron decay track are also merged. SV1 does not require any vertex merging. Only jets containing a single  $b$ -hadron at truth level are considered.

Next, vertices in the jet found by the different vertexing tools are compared with the target truth vertices. The number of correctly and incorrectly assigned tracks is computed. In order to call a vertex efficient, it is required to contain at least 65% of the tracks in the corresponding truth vertex, and to have a purity of at least 50%. Single track vertices are required to have a purity of 100%. Additionally, for GN1 only, at least one track in the vertex is required to have a predicted heavy flavour origin.

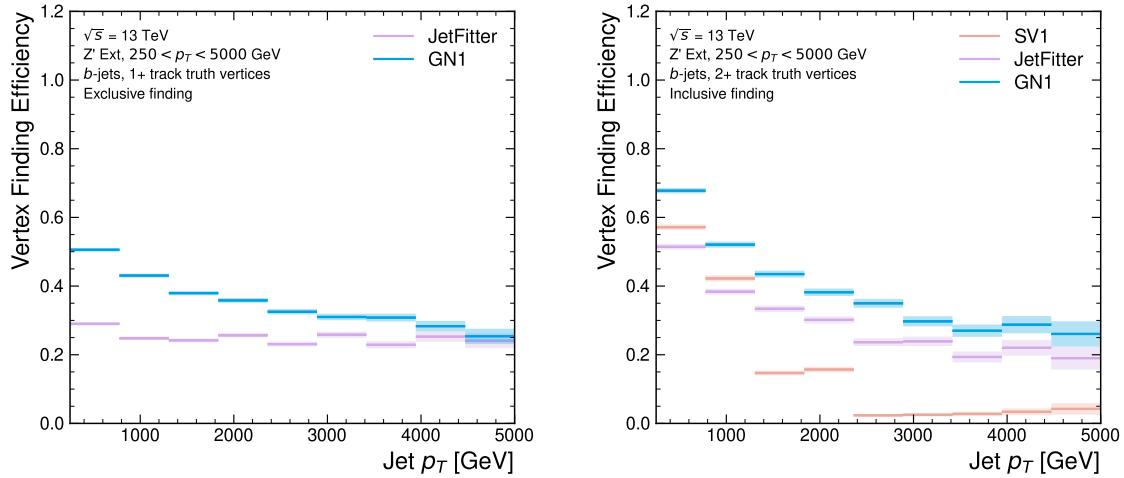
Vertex finding efficiencies for  $b$ -jets in the  $t\bar{t}$  sample are displayed as a function of  $p_T$  separately for the inclusive and exclusive approaches in Fig. 6.15. For  $b$ -jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV, the exclusive vertex finding efficiency of JetFitter and GN1 is relatively flat as a function of  $p_T$ . For the truth secondary vertices in this  $p_T$  region, JetFitter efficiently finds approximately 40% and GN1 finds approximately 55%. When finding vertices inclusively the vertex finding efficiency is generally higher. An increased dependence on  $p_T$  is also visible for JetFitter and SV1. As the jet  $p_T$  increases from 20 GeV to 100 GeV, the efficiency of JetFitter increases from 60% to 65%. In the same range, the efficiency of SV1 increases from 60% to 75%. GN1 displays less dependence on  $p_T$  than JetFitter and SV1, finding upwards of 80% of vertices in  $b$ -jets in this  $p_T$  region. For  $b$ -jets with  $p_T > 100$  GeV, JetFitter finds approximately 65% of vertices, SV1 finds approximately 75% of vertices, and GN1 finds approximately 80% of vertices.

Fig. 6.16 compares the exclusive and inclusive HF vertex finding efficiencies for  $b$ -jets in the  $Z'$  sample. The inclusive vertex finding efficiency drops steeply with increasing  $p_T$  up until  $p_T = 3$  TeV. GN1 outperforms SV1 and JetFitter across the  $p_T$  spectrum. In the first bin, the efficiency of GN1 is 65%, while the efficiencies of SV1 and JetFitter are around 55%. The efficiency of SV1 drops rapidly to almost zero above 3 TeV, while JetFitter and GN1 retain approximately 20% and 30% efficiency respectively. For the exclusive HF vertex finding efficiency, JetFitter finds 35% of vertices in the first bin, dropping to 20% of vertices above 2 TeV. GN1 finds 55% of vertices in the first bin, dropping to 30% above 2 TeV.

While Figs. 6.15 and 6.16 indicate that GN1 is able to successfully find displaced heavy flavour vertices in  $b$ -jets, it is also important to consider the vertexing performance inside light-jets. Light-jets may also contain real displaced vertices due to long lived secondary particles and material interactions. These tracks have a truth origin of OtherSecondary in the truth labelling scheme enumerated in see Table 5.1. The efficiency to reconstruct vertices comprised of OtherSecondary tracks can be computed in an analogous way to the heavy flavour vertexing efficiency, which is described above. Figs. 6.17 and 6.18 show the efficiency to reconstruct displaced OtherSecondary vertices in light-jets as a function of  $p_T$  for jets in the  $t\bar{t}$  sample and jets in the  $Z'$  sample respectively. The figures demonstrate that GN1 is able to more effectively find such vertices in light-jets as compared with SV1 and JetFitter. Since the properties of the displaced vertices in light-jets are likely to be significantly different to heavy

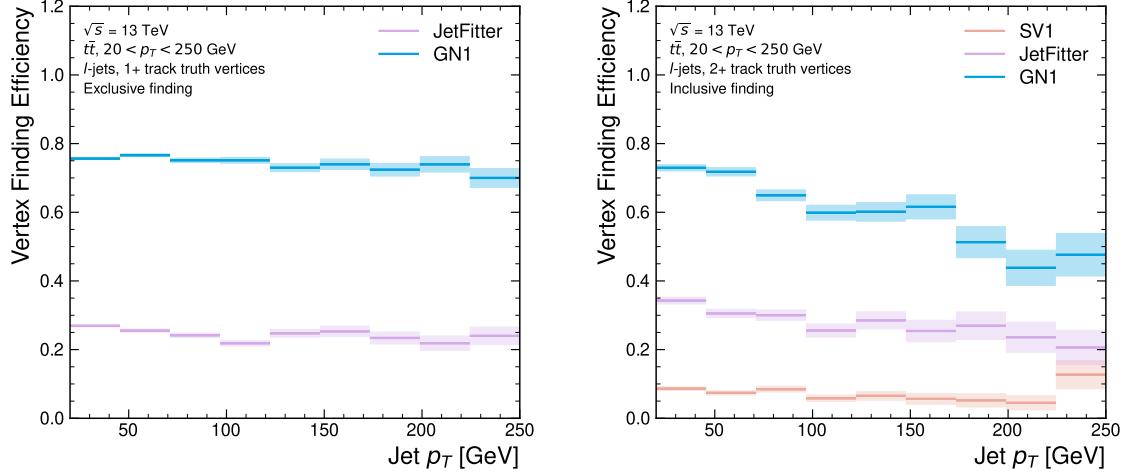


**Figure 6.15:** Heavy flavour vertex finding efficiency as a function of jet  $p_T$  for  $b$ -jets in the  $t\bar{t}$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.



**Figure 6.16:** Heavy flavour vertex finding efficiency as a function of jet  $p_T$  for  $b$ -jets in the  $Z'$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.

flavour vertices found in heavy flavour jets, the improved reconstruction of such vertices may help to differentiate between different flavour of jet.

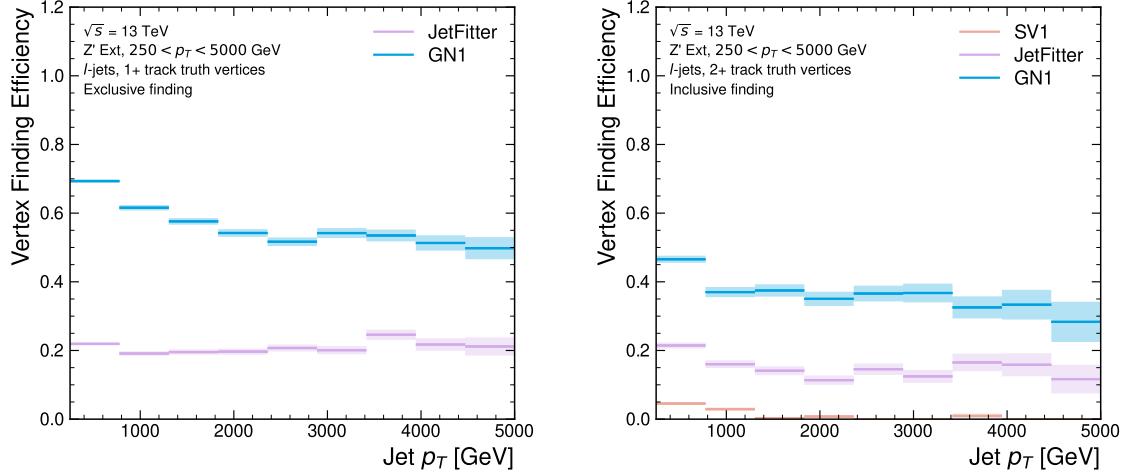


**Figure 6.17:** Vertex finding efficiency for other secondary decays as a function of jet  $p_T$  for light-jets in the  $t\bar{t}$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.

Collectively, the results in this section demonstrate that GN1 is able to accurately group tracks by their spatial origin in both  $b$ -jets and light-jets. The purity of the found vertices was also investigated and was found to be comparable or better than that of SV1 and JetFitter.

### 6.6.7 Track Classification Performance

One of the two auxiliary training objectives used by GN1 is to predict the truth origin of each track associated to the jet. Since the equivalent information is not provided by any of the existing flavour tagging tools, a benchmark model used to predict the truth origin of each track is trained based on a standard multi-class feed-forward classification network. The benchmark model is trained on the same tracks used for the baseline GN1 training. The model uses precisely the same concatenated track-and-jet inputs as used by GN1, but processes only a single track at a time, meaning



**Figure 6.18:** Vertex finding efficiency as a function of jet  $p_T$  for light-jets in the  $Z'$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.

it cannot take into account the correlations between tracks when determining the track origin. The model is made up of five densely connected linear layers with 200 neurons in each layer. The performance of the model was found to be insensitive to changes in the network structure.

To measure the track classification performance, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve is computed for each origin class, using a one-versus-all classification approach<sup>2</sup>. The AUCs for the different truth origins are averaged using both an unweighted and a weighted mean. The unweighted mean treats the performance of each class equally, while the weighted mean uses as a weight the relative abundance of tracks of each class. The same approach is used to compute the precision, recall and F1 scores<sup>3</sup>. Table 6.3 demonstrates clearly that GN1 outperforms the MLP both at  $20 < p_T < 250$  GeV for jets in the  $t\bar{t}$  sample and at  $250 < p_T < 5000$  GeV for jets in the  $Z'$  sample. For example, GN1 can reject

<sup>2</sup>One class is taken to be signal and the rest are taken as background, subsequently a binary classification approach is used.

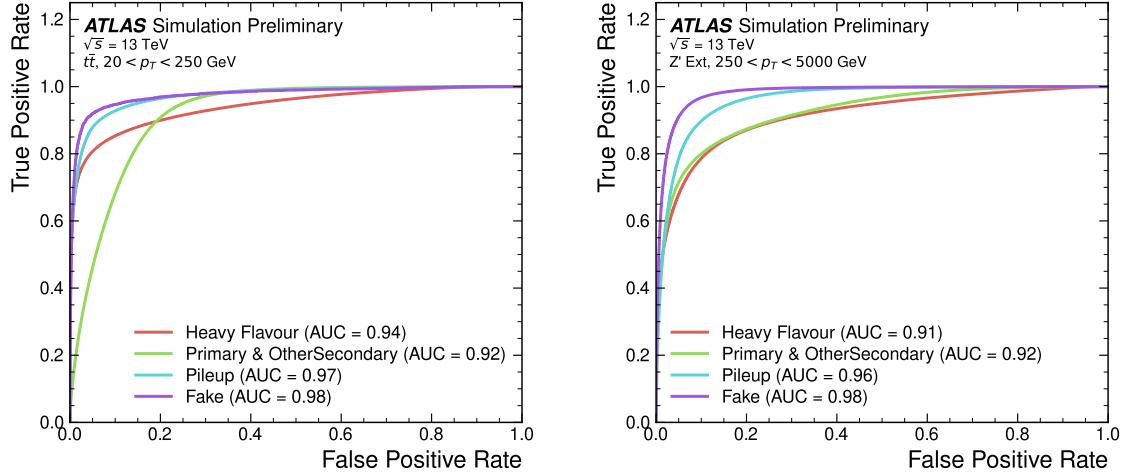
<sup>3</sup>The precision is the fraction of selected tracks from the signal class. The recall, similar to the efficiency, is the fraction of tracks from the signal class which are selected. The F1 score is the harmonic mean of the precision and recall.

65% of fake tracks in jets in the  $t\bar{t}$  sample, while retaining more than 99% of good tracks (i.e. those tracks which are not fake). The GN1 model has two advantages over the MLP which can explain the performance improvement. Firstly, the graph neural network architecture enables the sharing of information between tracks. This is likely to be beneficial since the origins of different tracks within a jet are correlated. Secondly, the jet classification and vertexing objectives may be complementary to the track classification objective, and so the track classification performance is improved by the combined training of complementary objectives.

		AUC		Precision		Recall		F1	
		Mean	Weighted	Mean	Weighted	Mean	Weighted	Mean	Weighted
$t\bar{t}$	MLP	0.87	0.89	0.39	0.71	0.51	0.56	0.36	0.62
	GN1	<b>0.92</b>	<b>0.95</b>	<b>0.51</b>	<b>0.82</b>	<b>0.64</b>	<b>0.70</b>	<b>0.51</b>	<b>0.74</b>
$Z'$	MLP	0.90	0.94	0.36	0.84	0.47	0.72	0.31	0.76
	GN1	<b>0.94</b>	<b>0.96</b>	<b>0.48</b>	<b>0.88</b>	<b>0.60</b>	<b>0.79</b>	<b>0.48</b>	<b>0.82</b>

**Table 6.3:** The area under the ROC curves (AUC) for the track classification from GN1, compared to a standard multilayer perceptron (MLP) trained on a per-track basis. The unweighted mean AUC over the origin classes and weighted mean AUC (using as a weight the fraction of tracks from the given origin) is provided. GN1, which uses an architecture that allows track origins to be classified in a conditional manner as discussed in Section 6.4.3, outperforms the MLP model for both  $t\bar{t}$  and  $Z'$  jets.

Fig. 6.19 shows the track origin classification ROC curves for the different track origins for jets in both the  $t\bar{t}$  and  $Z'$  samples. In order to improve visual readability of the plot, the curves for the heavy flavour truth origins (FromB, FromBC and FromC) have been combined (weighted by their relative abundance), as have the Primary and OtherSecondary origins. In jets in both the  $t\bar{t}$  and  $Z'$  samples, the AUC of all the different origin groups exceeds 0.9, representing strong overall classification performance. In both samples fake tracks are the easiest to classify, followed by pile-up tracks. The FromC tracks which are  $c$ -hadron decay products, are the hardest to classify, possibly due to their similarity to both fragmentation tracks and  $b$ -hadron decay tracks, depending on the  $c$ -hadron species in question.



**Figure 6.19:** ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the  $t\bar{t}$  sample (left) and jets in the  $Z'$  sample (right). The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the Primary and OtherSecondary labels have similarly been combined into a single category. The mean weighted area under the ROC curves (AUC) is similar for both samples [3].

## 6.7 Further Studies

### 6.7.1 Looser Track Selection

The track selections used to produce the main results are listed in Table 3.2. This includes a selection on the number of shared silicon modules used to reconstruct the track  $N_{\text{shared}}^{\text{Si}}$ . This value is calculated as

$$N_{\text{shared}}^{\text{Si}} = N_{\text{shared}}^{\text{Pix}} + N_{\text{shared}}^{\text{SCT}}/2 \quad (6.10)$$

where  $N_{\text{shared}}^{\text{Pix}}$  is the number of shared pixel hits and  $N_{\text{shared}}^{\text{SCT}}$  is the number of shared SCT modules on a track. The nominal selection used elsewhere in this thesis is  $N_{\text{shared}}^{\text{Si}} < 2$ . As the rate of shared hits is significantly higher for  $b$ -hadron decay tracks than for other tracks, especially at high- $p_T$ , this selection rejects a significant proportion of these tracks.

Figs. 6.20 and 6.21 show the result of training the GN1 tagger with the full relaxation of this selection, i.e. allowing tracks with any number of shared hits. The shared hit

requirements applied by the ambiguity solver as part of track reconstruction (see Section 3.4.1) are still applied. In addition, the maximum allowed value of  $d_0$  is increased from 3.5 mm to 5.0 mm. The results show that optimisation of the input track selection can lead to significant improvements in performance over the default selection. For the jets in the  $t\bar{t}$  sample shown in Fig. 6.20, the effect of loosening the track selection is limited. This is expected due to the lower prevalence of shared hits at lower transverse momenta. However for jets in the  $Z'$  sample as shown in Fig. 6.21, the  $c$ -jet rejection improves with respect to the baseline GN1 model by 35%, while the light-jet rejection improves by 40% at the 20%  $b$ -jet WP. Relative improvements of a similar magnitude are also observed for GN1Lep.

### 6.7.2 High Level Trigger

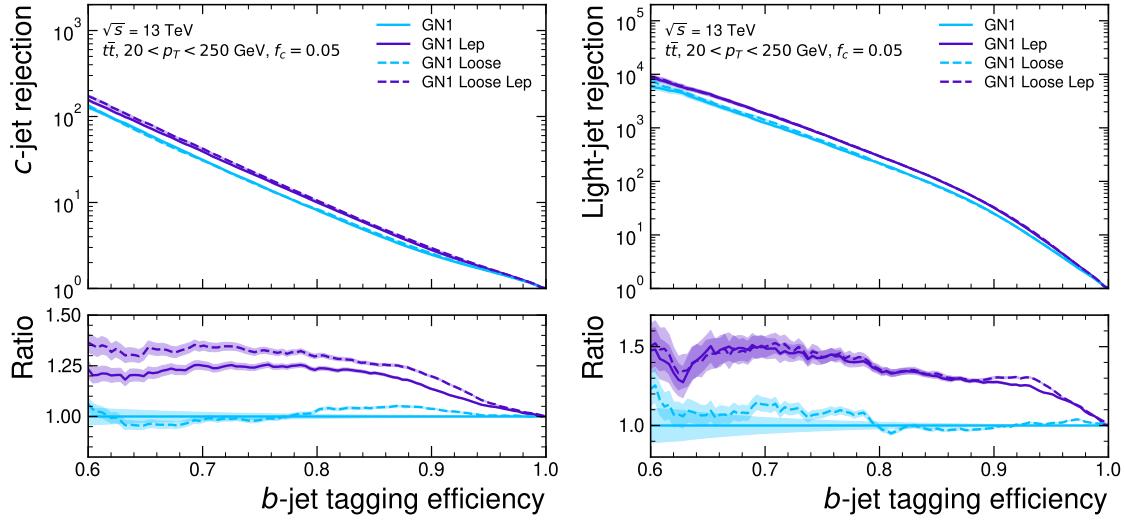
The implementation of GN1 described in this chapter has been re-used in several other contexts, demonstrating its flexibility to easily provide good jet flavour tagging performance with minimal overhead. The model has been implemented as a  $b$ -jet tagger in the High Level Trigger (HLT) (see Section 3.3.4). The inputs to the model are the running on precision tracks<sup>4</sup> and jet level quantities reconstructed after primary vertex reconstruction has been performed. Fig. 6.22 shows the performance of GN1 versus a comparable DL1d model [100], and two versions of DIPS [97], with EMTopo and PFlow jets (see Section 3.4.3) based on a low-precision region-of-interest based tracking pass, which is optimised for speed. The trigger implementation of GN1 improves upon the light-jet rejection of DL1d by 50% at the 60%  $b$ -jet WP for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV.

### 6.7.3 High Luminosity LHC

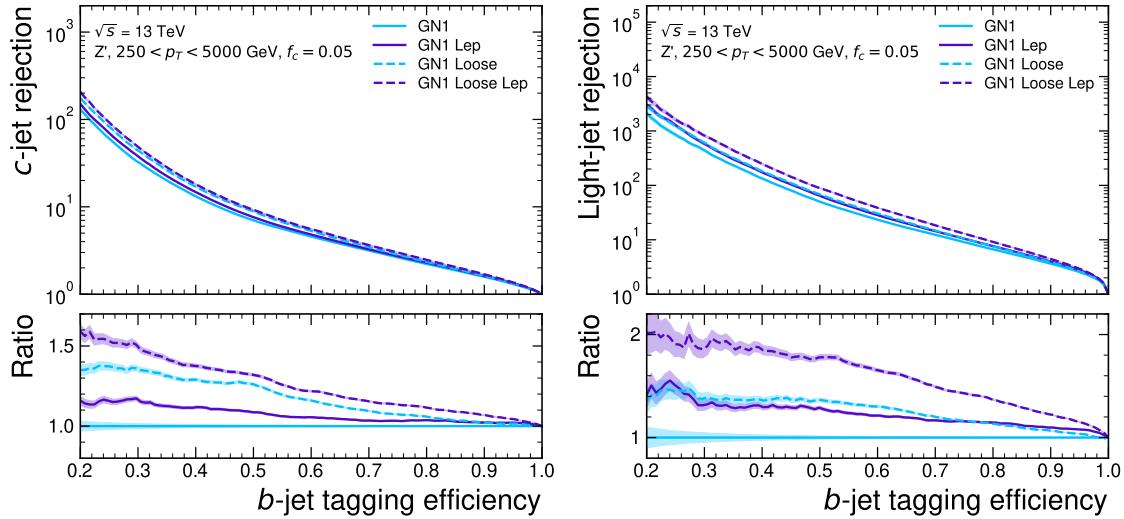
The model also demonstrates strong performance for the High Luminosity LHC (HL-LHC) using the proposed ITk inner tracking detector, as documented in Ref. [138]. Figs. 6.23 and 6.24 are reproduced from Ref. [138]. The results show that GN1 outperforms other existing flavour tagging algorithms when trained on an entirely different detector geometry, the ITk (see Section 3.3.1). When compared with

---

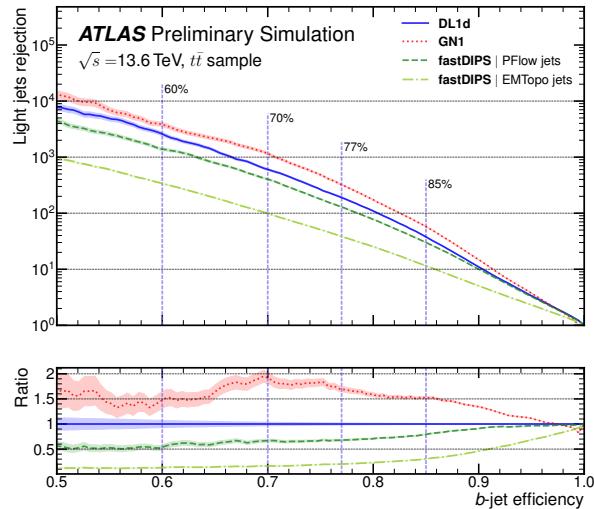
<sup>4</sup>Precision tracking refers to tracks reconstructed in the trigger using the same reconstruction algorithms as for the offline event reconstruction.



**Figure 6.20:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $t\bar{t}$  jets with  $20 < p_T < 250$  GeV, for the looser track selection trainings of GN1. The ratio to the performance of the baseline GN1 model is shown in the bottom panels. A value of  $f_c = 0.05$  is used for all models. Binomial error bands are denoted by the shaded regions.



**Figure 6.21:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $Z'$  jets with  $250 < p_T < 5000$  GeV, for the looser track selection trainings of GN1. The ratio to the performance of the baseline GN1 model is shown in the bottom panels. A value of  $f_c = 0.05$  is used for all models. Binomial error bands are denoted by the shaded regions.



**Figure 6.22:** The HLT light-jet rejection as a function of the  $b$ -jet efficiency jets in the  $t\bar{t}$  sample with  $20 < p_T < 250 \text{ GeV}$  for events with a centre of mass energy  $\sqrt{s} = 13.6 \text{ TeV}$  [137]. The ratio to the performance of the DL1d algorithm [100] is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for b-tagging.

DL1d [100], GN1 improves on the  $c$ -jet rejection (light-jet rejection) by a factor of  $\sim 2$  ( $\sim 2.5$ ) for jets in the  $t\bar{t}$  sample at the 60%  $b$ -jet WP. Significant improvements in rejections are also observed for jets in the  $Z'$  sample.

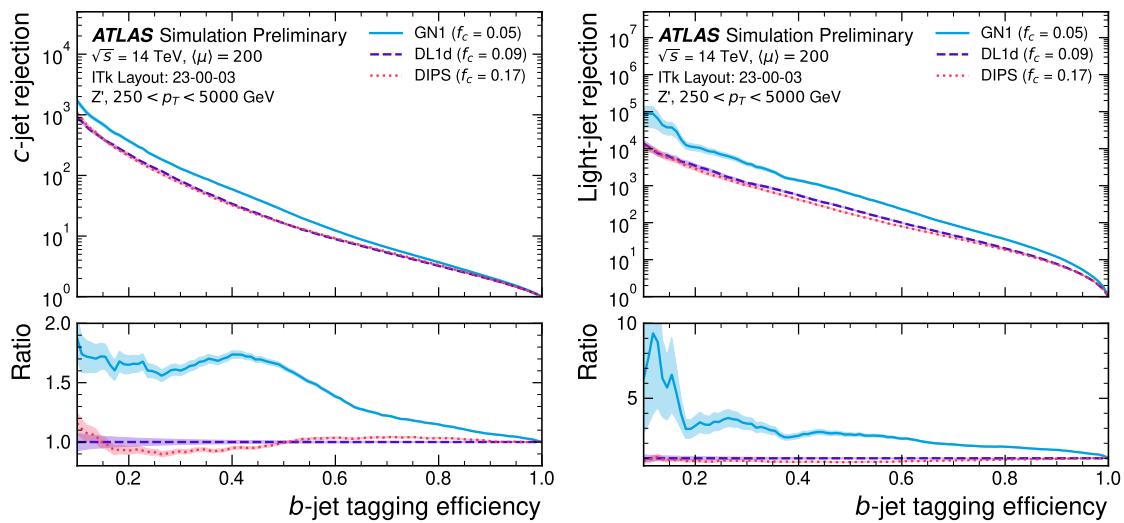
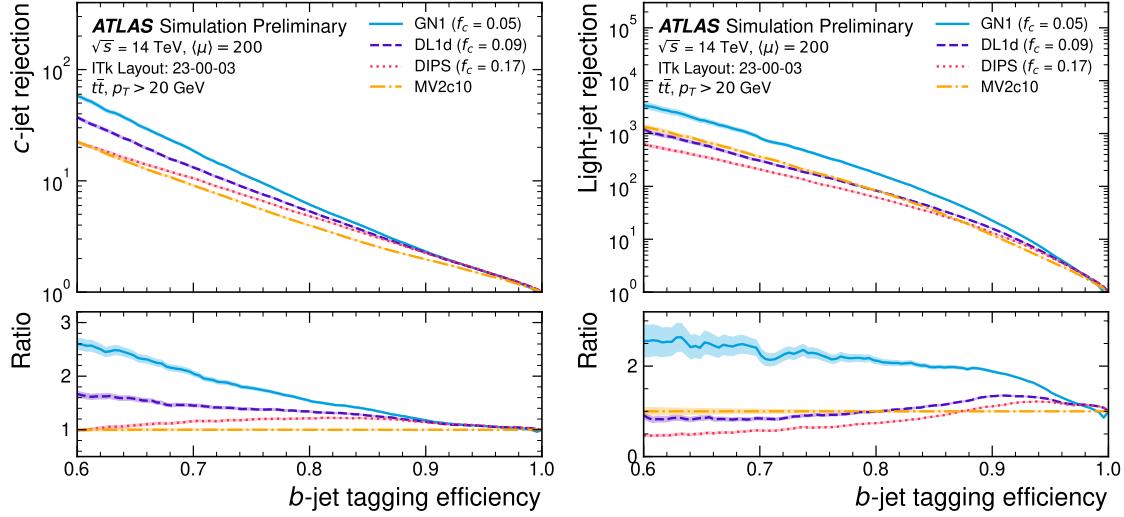
## 6.8 Conclusion

In this chapter a novel jet tagger, GN1, is presented. The model has a graph neural network architecture and is trained with auxiliary training objectives, which are shown to improve the performance of the basic model. GN1 significantly improves flavour tagging performance with respect to DL1r, the current default ATLAS flavour tagging algorithm, when compared in simulated collisions. GN1 improves  $c$ - and light-jet rejection for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV by factors of  $\sim 2.1$  and  $\sim 1.8$  respectively at a  $b$ -jet tagging efficiency of 70% when compared with DL1r. For jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV, GN1 improves the  $c$ -jet rejection by a factor of  $\sim 2.8$  and light-jet rejection by a factor of  $\sim 6$  for a comparative  $b$ -jet efficiency of 30%.

Previous multivariate flavour tagging algorithms relied on inputs from low-level tagging algorithms, whereas GN1 needs no such inputs, making it more flexible. It can be easily fully optimised via a retraining for specific flavour tagging use cases, as demonstrated with  $c$ -tagging and high- $p_T$   $b$ -tagging, without the need for time-consuming retuning of the low-level tagging algorithms. The model is also simpler to maintain and study due to the reduction in the number of constituent components.

GN1 demonstrates improved track classification performance when compared with a simple per-track MLP. The model is also able to perform vertex finding, and preliminary studies suggest it outperforms previous manually optimised approaches. The auxiliary track classification and vertex finding objectives are shown to significantly contribute to the performance in the jet classification objective, and, along with the more advanced graph neural network architecture, are directly responsible for the improvement over DL1r.

GN1 has also been shown to perform well for  $b$ -tagging in the High Level Trigger, and in the projected environment for the upgraded High Luminosity LHC. The



performance of the model at high- $p_{\mathrm{T}}$  can be improved by loosening the input track selection, which increases the number of  $b$ -hadron decay tracks that are inputted to the model.

Preliminary validation of the model demonstrates that the level of discrepancy between different Monte Carlo event generators is similar to that seen for previous ATLAS ML based taggers such as DL1r, suggesting that the model has not learnt additional information that is specific to a single event generator. Initial comparisons between simulated events and data also show similar levels of agreement as for previous taggers [139].

## Future Work

Further improvements in the  $b$ - and  $c$ -tagging performance are likely possible with a more thorough optimisation of the model architecture, and the integration of additional information from other parts of the ATLAS detector. The addition of other auxiliary training objectives, such as the prediction of the truth  $b$ -hadron decay radius and transverse momentum and the truth type of input tracks may also yield additional performance gains.

Although the results demonstrate a significant performance improvement at high- $p_{\mathrm{T}}$ , it is also possible that additional studies on further loosening the selection could yield further improved results. For example the selections on the number of number of holes and the longitudinal impact parameter could be further relaxed. The maximum number of tracks provided as input to the model could also be increased from the default value of 40. In order to validate the change from the default tracking setup, studies investigating the modelling uncertainties of the additional tracks need to be carried out.

Given that GN1 exploits the input information in a more sophisticated way than DL1r, further studies are needed to confirm that the performance gain observed in these simulated samples is also observed in experimental data. Additional future work includes the full calibration of the model so it can be used by physics analyses.

The flexible nature of the model means it can also be readily applied to other related problems outside of standard  $b$ - and  $c$ -tagging applications, as demonstrated in Section 6.7. Additional applications for the architecture include  $X \rightarrow bb$  and  $X \rightarrow cc$

tagging. The model could also be repurposed as a pile-up jet or  $\tau$  tagger, or general primary and secondary vertexing tool. Using a vertex fitting algorithm to compare the reconstructed vertex quantities with those from SV1 and JetFitter is left for future work.

# Chapter 7

## Boosted VHbb Analysis

The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [16, 17], is predicted by the Standard Model to decay primarily to a pair of  $b$ -quarks, with a branching fraction of  $0.582 \pm 0.007$  for  $m_H = 125$  GeV [28]. Observation of this decay mode was reported by ATLAS [89] and CMS [30] in 2018, establishing the first direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks (see Section 2.2.2). The  $H \rightarrow b\bar{b}$  process is also important for constraining the total decay width of the Higgs [140].

Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet background and so overall sensitivity to the Higgs is low. The QCD multijet background refers to events containing one or more strongly produced jets which are not the decay product of heavy resonances, for example  $g \rightarrow q\bar{q}$ . In the  $H \rightarrow b\bar{b}$  gluon-gluon fusion channel the vast majority of events will contain only jets in the final state, and therefore it is extremely difficult to distinguish signal events from the overwhelming multijet background. The  $H \rightarrow b\bar{b}$  observation therefore searched for Higgs bosons produced in association with a vector boson  $V$  (where  $V$  can be a  $W$  or  $Z$  boson) which subsequently decays leptonically. The leptonic final states allow for leptonic triggering whilst significantly reducing the multijet background.

Two full Run 2 dataset analyses were carried out as a follow-up to the  $H \rightarrow b\bar{b}$  observation [89]. Similar to the observation, both measured the associated production of a Higgs boson with a vector boson, with the Higgs boson decaying to a pair of  $b$ -quarks. The first analysis [141] was focused on the resolved phase-space, where the Higgs candidate is reconstructed as two distinct jets with radius parameter

$R = 0.4$ . The second analysis [142] was focused on the boosted phase-space, where the Higgs candidate had a sufficiently large transverse momenta such that it can be reconstructed as a single jet with a radius parameter of  $R = 1.0$ . This chapter will focus on the latter analysis, which is referred to as the boosted  $VH$ ,  $H \rightarrow b\bar{b}$  analysis.

In this chapter, the boosted  $VH$ ,  $H \rightarrow b\bar{b}$  analysis is outlined in Section 7.1. An introduction to the systematic uncertainties used in the analysis is provided in Section 7.1.6. Section 7.2 briefly outlines the experiemntal uncertainties impacting the analysis. Detailed modelling studies performed for the analysis are described in Section 7.3. The statistical treatment is detailed in Section 7.4, and the results of the analysis are presented in Section 7.5. This analysis has been published in Ref. [142]. Figures and tables from the published work are reproduced here.

## 7.1 Analysis Overview

The boosted  $VH$ ,  $H \rightarrow b\bar{b}$  analysis is focused on the high transverse momentum regime, which has the benefit of being more sensitive to physics beyond the Standard Model [143], but the disadvantage of being more challenging due to the increased difficulty in the accurate reconstruction of high transverse momentum physics objects (discussed in Chapter 4). In order to focus on the high- $p_T$  regime, the reconstructed vector boson  $p_T^V$  is required to be  $p_T^V > 250 \text{ GeV}$  (see Section 7.1.1). Events are also split into two  $p_T$  bins with the first bin covering  $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$  and the second covering  $p_T^V > 400 \text{ GeV}$ , which allows the analysis to benefit from the improved signal-to-background ratio in the high- $p_T$  regime [144].

The previous ATLAS analysis in Ref. [89] was primarily sensitive to vector bosons with a more modest  $p_T^V$  boost in the region of 150–300 GeV (for the 0- and 1-lepton channels) and 75–300 GeV (for the 2-lepton channel). In this regime, the Higgs candidate was reconstructed using a pair of jets with radius parameter of  $R = 0.4$ , called small- $R$  jets. However in the high- $p_T$  regime, the decay products of the Higgs boson become increasingly collimated and the small- $R$  jets may not be individually resolved. In order to enhance the reconstruction of the Higgs boson candidate, this analysis uses a large- $R$  jet with radius parameter  $R = 1.0$  to reconstruct the Higgs boson candidate. The Higgs candidate is required to have exactly two ghost-

associated (see Section 3.4.3) and  $b$ -tagged variable-radius track-jets ( $b$ -track-jets). The candidate large- $R$  jet is reconstructed using jet substructure techniques, in particular it is trimmed by removing soft and wide-angle components, which helps to remove particles from the underlying event and pile-up collisions [79]. Refer to Section 3.4.3 for more details on jet reconstruction.

A complete overview of the different analysis regions is given in Table 7.1. On top of the binning in  $p_T^V$ , selected events are further categorised into the 0-, 1- and 2-lepton channels depending on the number of charged leptons (electrons and muons) present in the reconstructed final state (also referred to as the 0L, 1L, and 2L channels respectively). The 0-lepton channel targets the  $ZH \rightarrow \nu\nu b\bar{b}$  process, the 1-lepton channel targets  $WH \rightarrow \ell\nu b\bar{b}$ , and the 2-lepton channel targets  $ZH \rightarrow \ell\ell b\bar{b}$ , where  $\ell$  is an electron or muon and  $\nu$  is a neutrino. Each channel has a dedicated set of selections which are listed in more detail in Section 7.1.2. Events in the 0- and 1-lepton channels are further split depending on the number of additional small- $R$  jets not matched to the Higgs candidate. The high-purity signal region (HP SR) has zero such jets, while the low-purity signal region (LP SR) has one or more, and therefore absorbs a larger number of background  $t\bar{t}$  events. Maintaining a high purity signal region is important for the extraction of the signal yield. The 0- and 1-lepton channels also make use of a dedicated  $t\bar{t}$  control region for events with one or more additional  $b$ -tagged small- $R$  jets outside the large- $R$  jet, as described in Section 7.1.3.

The large- $R$  jet mass is used as the main discriminant in the analysis. The  $VH$ ,  $H \rightarrow b\bar{b}$  signal yield is extracted from a profile likelihood fit to the large- $R$  jet mass over several signal and control analysis regions, which are described in Sections 7.1.2 and 7.1.3 respectively. The diboson background  $VZ$ ,  $Z \rightarrow b\bar{b}$  yield is simultaneously extracted from the fit, and provides a cross check on the  $VH$  signal extraction. The statistical fit model (described henceforth only as “the fit”) is described in more detail in described in Section 7.4.

### 7.1.1 Object Reconstruction

The 0- and 1-lepton channels make use of reconstructed missing transverse energy (see Section 3.4.5). The presence of neutrinos in the  $ZH \rightarrow \nu\nu b\bar{b}$  and  $WH \rightarrow \ell\nu b\bar{b}$  signatures can be inferred from a momentum imbalance in the transverse plane. The

		Analysis Regions					
		$250 < p_T^V < 400 \text{ GeV}$		$p_T^V \geq 400 \text{ GeV}$			
Channel	0 add. $b$ -track-jets	$\geq 1$ add. $b$ -track-jets	0 add. $b$ -track-jets	$\geq 1$ add. $b$ -track-jets			
	0 add. small- $R$ jets	$\geq 1$ add. small- $R$ jets	0 add. small- $R$ jets	$\geq 1$ add. small- $R$ jets			
0-lepton	HP SR	LP SR	CR	HP SR	LP SR	SR	CR
1-lepton	HP SR	LP SR	CR	HP SR	LP SR	SR	CR
2-lepton	SR			SR			

**Table 7.1:** Summary of the definitions of the different analysis regions . Signal enriched regions are marked with the label SR. There are regions with relatively large signal purity (HP SR) and with low purity (LP SR). Background enriched control regions are marked with the label CR. The shorthand “add” stands for additional small- $R$  jets, i.e. number of small- $R$  jets not matched to the Higgs-jet candidate. The medium and high  $p_T^V$  regions are referred to as  $Mp_T^V$  and  $Hp_T^V$ , respectively [142].

vector boson transverse momentum  $p_T^V$  is reconstructed as the missing transverse energy  $E_T^{\text{miss}}$  in the 0-lepton channel, as the magnitude of the summed  $\mathbf{E}_T^{\text{miss}}$  and charged-lepton momentum in the 1-lepton channel, and as the transverse momentum of the 2-lepton system in the 2-lepton channel.

Electrons and muons are reconstructed and identified as outlined in Section 3.4.4, and following the approach described in Ref. [89]. Leptons are required to satisfy the selections listed in Table 7.2. *Baseline* electrons are required to pass the looser likelihood-based identification selection described in Section 3.4.4, whilst *Signal* electrons are required to satisfy a tighter likelihood identification selection. *Baseline* muons are required to pass the ‘loose’ identification described in Ref. [56], while *signal* muons are required to pass the ‘medium’ identification working point. All signal leptons are required to additionally satisfy a  $p_T > 27 \text{ GeV}$  selection criteria, except for muons in the 1-lepton channel where a cut of 25 GeV is used. The number of baseline leptons is used to categorise the event into the 0-, 1- or 2-lepton channels. The 1- and 2-lepton channels require one signal lepton to be present.

The analysis makes use of large- $R$  and small- $R$  EMTopo jets, and also variable-radius track-jets. The different types of jets are outlined in Section 3.4.3. The large- $R$  jets are required to satisfy  $p_T > 250 \text{ GeV}$  and  $|\eta| < 2.0$ , and are used to reconstruct the Higgs boson candidate, while small- $R$  jets are used in the selection of the analysis

Variable	Electrons	Muons
$p_T$	$> 7 \text{ GeV}$	
$ \eta $	$< 2.47$	$< 2.7$
$s(d_0)$	$< 5$	$< 3$
$ z_0 \sin(\theta) $	$< 0.5 \text{ mm}$	

**Table 7.2:** Selections applied to baseline and signal electrons and muons. Signal leptons are additionally required to satisfy a  $p_T > 27 \text{ GeV}$  ( $25 \text{ GeV}$  in the 1-lepton channel).

region and in the construction of  $E_T^{\text{miss}}$ . The track-jets are used for  $b$ -tagging and are required to have at least two constituent tracks with  $p_T > 500 \text{ MeV}$  and  $|\eta| < 2.5$ , and must themselves satisfy  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.5$ . The track-jets matched to the Higgs candidate are  $b$ -tagged using the MV2c10  $b$ -tagging algorithm (see Chapter 4). The efficiency of the tagging algorithm is calibrated to events in data [145–147]. The jet tagging strategy relies on extensive studies into track-jet  $b$ -tagging in boosted topologies [148, 149].

### 7.1.2 Event Selection

An extensive list of selection cuts are applied to each event in order to reject background events whilst retaining as many signal events as possible. A full list of selection cuts applied to the different analysis regions is given in Table 7.3, while some key selections are listed below.

All channels require events with at least one large- $R$  jet. The vector boson transverse momentum is also required to satisfy  $p_T^V > 250 \text{ GeV}$ . The Higgs candidate is chosen as the highest  $p_T$  large- $R$  jet satisfying these requirements. As mentioned, the candidate large- $R$  jet is required to have two ghost-associated and  $b$ -tagged variable-radius track-jets.

In the 0-lepton channel, trigger selections are applied using an  $E_T^{\text{miss}}$  trigger with a luminosity-dependent threshold that varied between  $70\text{--}110 \text{ GeV}$  depending on the data taking period. In the 1-lepton electron sub-channel a combination of single electron triggers is used with minimum  $p_T$  thresholds between  $24\text{--}26 \text{ GeV}$ . In the muon sub-channel the same  $E_T^{\text{miss}}$  trigger as the 0-lepton channel is used. Since muons

are not used for the  $E_T^{\text{miss}}$  trigger calculations, this is in effect a  $p_T$  requirement on the muon-neutrino system, which in the analysis phase space is more efficient than a single-muon trigger. The 2-lepton channel uses the same triggering strategy as the 1-lepton channel. In all channels, the trigger selections applied are fully efficient for events selected using the full requirements in Table 7.3.

The combined selections in Table 7.3 result in a signal efficiency ranging from 6–16% for the  $WH$  and  $ZH$  processes depending on the channel and  $p_T^V$  bin.

### 7.1.3 Control Regions

The  $t\bar{t}$  process presents a major background in the 0- and 1-lepton channels. In these events, the Higgs candidate is often reconstructed from a correctly tagged  $b$ -jet from the top decay  $t \rightarrow Wb$ , and an incorrectly tagged  $c$ - or light-jet from the subsequent decay of the  $W$ , as shown in Fig. 7.1.

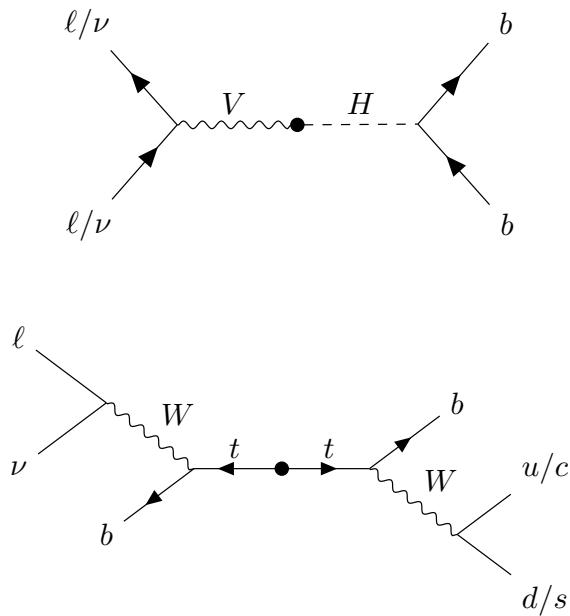
The top quark decays to a  $W$  and a  $b$ -quark the vast majority of the time. Hence, the second top quark from the  $t\bar{t}$  pair is also likely to result in a second tagged  $b$ -tagged track-jet outside of the large- $R$  Higgs candidate. To ensure sufficient  $t\bar{t}$  rejection, 0- and 1-lepton channel signal regions are defined using a veto on events with  $b$ -tagged track-jets outside the Higgs candidate. These events are used to construct a control region (CR) which is enriched in  $t\bar{t}$  events. The CR is used to constrain the normalisation of the  $t\bar{t}$  background in the fit.

### 7.1.4 Background Composition

After the selections described in Section 7.1.2 the number of background events mimicking the  $VH$ ,  $H \rightarrow b\bar{b}$  signal is greatly reduced. However, the number of background events still greatly outnumbers that of signal events. The background processes are channel dependent. In the 0-lepton channel the dominant sources of backgrounds are  $Z+\text{jets}$  ( $Z \rightarrow \nu\nu$ ) and  $t\bar{t}$ , with  $W+\text{jets}$  and diboson events being subdominant.  $W+\text{jets}$  contribute in the event of  $W \rightarrow \tau\nu$ , and subsequent hadronic decay of the  $\tau$  or lack of successful reconstruction/selection of the leptonic decay products.  $t\bar{t}$  and  $W+\text{jets}$  (with a leptonic decay of the  $W$  as in  $W \rightarrow \ell\nu$ ) are

Selection	0 lepton channel	1 lepton channel	2 leptons channel	
	$e$ sub-channel	$\mu$ sub-channel	$e$ sub-channel	$\mu$ sub-channel
Trigger	$E_T^{\text{miss}}$	Single electron	$E_T^{\text{miss}}$	$E_T^{\text{miss}}$
Leptons	0 baseline leptons	1 signal lepton $p_T > 27 \text{ GeV}$	$p_T > 25 \text{ GeV}$	2 baseline leptons among which $\geq 1$ signal lepton, $p_T > 27 \text{ GeV}$ both leptons of the same flavour
		no second baseline lepton	-	-
				opposite sign muons
$E_T^{\text{miss}}$	$> 250 \text{ GeV}$	$> 50 \text{ GeV}$	-	-
$p_T^V$			$p_T^V > 250 \text{ GeV}$	
Large- $R$ jets		at least one large- $R$ jet, $p_T > 250 \text{ GeV},  \eta  < 2.0$		
Track-jets		at least two track-jets, $p_T > 10 \text{ GeV},  \eta  < 2.5$ , matched to the leading large- $R$ jet		
$b$ -tagged jets		leading two track-jets matched to the leading large- $R$ must be $b$ -tagged (MV2c10, 70%)		
$m_J$			$> 50 \text{ GeV}$	
$\min[\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{small-}R \text{ jets})]$	$> 30^\circ$		-	
$\Delta\phi(\mathbf{E}_T^{\text{miss}}, H_{\text{cand}})$	$> 120^\circ$		-	
$\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T,\text{trk}}^{\text{miss}})$	$< 90^\circ$		-	
$\Delta y(V, H_{\text{cand}})$	-		$ \Delta y(V, H_{\text{cand}})  < 1.4$	
$m_{\ell\ell}$				$66 \text{ GeV} < m_{\ell\ell} < 116 \text{ GeV}$
Lepton $p_T$ imbalance		-		$(p_T^{\ell_1} - p_T^{\ell_2})/p_T^Z < 0.8$

**Table 7.3:** Event selection requirements for the boosted  $VH$ ,  $H \rightarrow b\bar{b}$  analysis channels and sub-channels [142]. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region. The  $\min[\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{small-}R \text{ jets})]$  selection is used to remove jets when the missing transverse momentum  $\mathbf{E}_T^{\text{miss}}$  is pointing in the direction of the Higgs candidate, and the  $\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T,\text{trk}}^{\text{miss}})$  is used to reject events where the calorimeter missing transverse momentum  $\mathbf{E}_T^{\text{miss}}$  is not pointing in the direction of the track-based missing transverse momentum  $\mathbf{E}_{T,\text{trk}}^{\text{miss}}$ . The  $\Delta y(V, H_{\text{cand}})$  quantifies the rapidity difference between the reconstructed vector boson and Higgs candidate ( $H_{\text{cand}}$ ).  $m_{\ell\ell}$  refers to the invariant mass of the two leptons, while  $p_T^{\ell_1}$  and  $p_T^{\ell_2}$  refer to the  $p_T$  of the first and second leptons, respectively.



**Figure 7.1:** Diagrams of the signal process (top) and  $t\bar{t}$  background (bottom). The object to the right of centre are reconstructed within the large- $R$  jet. For the  $t\bar{t}$  background, the large- $R$  jet contains a mistagged  $c$ -jet or (less often) a mistagged light-jet. The contribution in the 0-lepton channel results from hadronically decaying  $\tau$  lepton, or a electron or muon which is out of the analysis acceptance. For the  $t\bar{t}$  background process to be in the signal region, the  $b$ -jet on the left hand side must either be not within acceptance, not reconstructed, or not tagged.

dominant in the 1-lepton channel, while single-top is subdominant. In the 2-lepton channel,  $Z + \text{jets}$  ( $Z \rightarrow \ell\ell$ ) is again dominant followed by  $ZZ$  diboson events.

The diboson background  $VV$  consists primarily of  $WZ$  and  $ZZ$  events in which the  $Z$  decays to a pair of  $b$ -quarks. This process very closely matches the signal, with a resonant peak occurring at  $m_Z = 91$  GeV. As it is a well understood process, the diboson background can be used as a cross-check on the main  $H \rightarrow b\bar{b}$  results.

The  $t\bar{t}V$ ,  $t\bar{t}H$  and multijet backgrounds are negligible in the analysis phase space after the selections have been applied, with the exception of the 1-lepton electron sub-channel, in which multijet background is not negligible. The multijet background in this region is generally made up of events with mismeasured jets and events where the isolated leptonic signature has been mimicked by either a jet or electron from a semi-leptonic heavy flavour decay, where the electron has escaped the jet.

The contributions from the different backgrounds are modelled using Monte Carlo event generators and the uncertainties associated with these samples are studied in Section 7.3. The multijet background is modelled using a data-driven technique.

### 7.1.5 Data & Simulated Samples

Data from centre-of-mass energy  $\sqrt{s} = 13$  TeV proton-proton collisions at the LHC recorded over the course of Run 2 (between 2015 and 2018) were used for the analysis. The resulting dataset corresponds to a total integrated luminosity of  $139\text{ fb}^{-1}$ .

An overview of the MC simulated samples used in the analysis is given in Table 7.4. These samples are used to model the signal and background processes relevant to the analysis, with the exception of the multijet background which is modelled using a data-driven technique. Data and simulated events are reconstructed using the same algorithms, and a reweighting is applied to the simulated events in order to match the pile-up distribution observed in the data.

### 7.1.6 Overview of Systematic Uncertainties

Systemic uncertainties are extensively employed to reflect assumptions and inaccuracies in the various inputs used by the fit. Two main types of systematic uncertainty

Process	ME generator	ME PDF	PS and Hadronisation	UE model tune	Cross-section order
Signal ( $m_H = 125$ GeV and $b\bar{b}$ branching fraction set to 58%)					
$gg \rightarrow WH \rightarrow \ell\nu b\bar{b}$	Powheg-Box v2 [150] + GoSAM [152] + MiNLO [153, 154]	NNPDF3.0NLO (*) [105]	Pythia 8.212 [107]	AZNLO [151]	NNLO(QCD) + NLO(EW) [155–161]
$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg-Box v2 + GoSAM + MiNLO	NNPDF3.0NLO (*)	Pythia 8.212	AZNLO	NNLO(QCD) <sup>(†)</sup> + NLO(EW)
$gg \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$	Powheg-Box v2	NNPDF3.0NLO (*)	Pythia 8.212	AZNLO	NLO + NLL [162–166]
Top quark ( $m_t = 172.5$ GeV)					
$t\bar{t}$	Powheg-Box v2 [150, 167]	NNPDF3.0NLO	Pythia 8.230	A14 [109]	NNLO+NNLL [168]
s-channel	Powheg-Box v2 [150, 169]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [170]
t-channel	Powheg-Box v2 [150, 169]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [171]
$Wt$	Powheg-Box v2 [150, 172]	NNPDF3.0NLO	Pythia 8.230	A14	Approximate NNLO [173]
Vector boson + jets					
$W \rightarrow \ell\nu$	SHERPA 2.2.1 [174–177]	NNPDF3.0NNLO	SHERPA 2.2.1 [178, 179]	Default	NNLO [180]
$Z/\gamma^* \rightarrow \ell\ell$	SHERPA 2.2.1	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NNLO
$Z \rightarrow \nu\nu$	SHERPA 2.2.1	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NNLO
Diboson					
$qq \rightarrow WW$	SHERPA 2.2.1	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NLO
$qq \rightarrow WZ$	SHERPA 2.2.1	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NLO
$qq \rightarrow ZZ$	SHERPA 2.2.1	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NLO
$gg \rightarrow VV$	SHERPA 2.2.2	NNPDF3.0NNLO	SHERPA 2.2.2	Default	NLO

**Table 7.4:** Signal and background processes with the corresponding generators used for the nominal samples [142]. If not specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant ( $\alpha_s$ ). (\*) The events were generated using the first PDF in the NNPDF3.0NLO set and subsequently reweighted to the PDF4LHC15NLO set [181] using the internal algorithm in POWHEG-BOX v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for the  $pp \rightarrow ZH$  process already includes the  $gg \rightarrow ZH$  contribution. The  $gg \rightarrow ZH$  process is normalised using the cross-section for the  $pp \rightarrow ZH$  process, after subtracting the  $gg \rightarrow ZH$  contribution. An additional scale factor is applied to the  $gg \rightarrow VH$  processes as a function of the transverse momentum of the vector boson, to account for electroweak (EW) corrections at NLO. This makes use of the  $VH$  differential cross-section computed with HAWK [182, 183].

are considered: experimental and modelling. Experimental uncertainties arise due to the imperfect modelling of the reconstruction algorithms (in particular the jet reconstruction and  $b$ -tagging algorithms), and due to the imperfect modelling of pile-up and other effects, as described in Section 7.2. Modelling uncertainties are described in Section 7.3 and arise due to the imperfections in the Monte-Carlo generators used to model the signal and background events. In order to observe a certain process, for example  $VH$ ,  $H \rightarrow b\bar{b}$ , an increase in the number of observed events with respect to the background-only hypothesis is looked for. The excess is often relatively small against the total number of background events, and hence accurate modelling of the expected number of background and signal events is crucial for successfully performing the analysis.

## 7.2 Experimental Uncertainties

The main experimental uncertainties in the analysis are due to the modelling of following sources:

- The small- $R$  jet energy scale and resolution, which are informed by in situ calibration studies [75].
- The large- $R$  jet energy and mass scales and resolutions. The scales are calibrated as described in Ref. [81], and an uncertainty of 2% and 20% is applied for the jet energy and mass resolutions, respectively [184, 185].
- $b$ -tagging uncertainties, which are computed separately for  $b$ -,  $c$ - and light-flavour jets as described in the data-based calibration studies in Refs. [145–147]. An additional extrapolation uncertainty is added to account for jets with transverse momenta above that which is accessible in the calibration analyses.
- Uncertainties associated with the lepton energy and momentum scales, and reconstruction and identification efficiencies.
- Uncertainty on the pile-up models which are used in the simulated samples, described in Ref. [186].
- Uncertainties associated with the reconstruction of the missing transverse energy  $E_T^{\text{miss}}$ , as described in Ref. [88].

- Uncertainty on the total recorded integrated luminosity, as described in Ref. [42].

The impact of these uncertainties on the analysis can be found in Table 7.18.

## 7.3 Modelling Uncertainties

Particular care is paid to the uncertainties on the modelling predictions as discussed in this section. The *Nominal* MC samples are used as a reference to which different variations can be compared. The nominal samples are chosen as the best possible representation of the underlying physical process. *Alternative* samples are used to understand inaccuracies that may be present in the nominal samples. Some aspect of the nominal model is varied, and the discrepancy with respect to the nominal model is quantified. The discrepancy is used to estimate a systematic uncertainty associated with the model parameter which was varied. The alternate samples are sometimes obtained via internal weight variations or parameterisation methods, rather than by re-running the simulation. This is discussed in more detail in Section 7.3.3.

Modelling systematics can have several impacts, including affecting the overall normalisation for different processes, the relative acceptances between different analysis regions (i.e. migrations between HP and LP SRs, between the SR and CR, and between  $p_T^V$  bins), and the shapes of the  $m_j$  distributions. For each source of uncertainty, acceptance and shape uncertainties are therefore derived. Acceptance uncertainties account for the uncertainty in the overall number of events in each channel, and for the relative numbers of events in the various analysis regions. Meanwhile, shape uncertainties account for the uncertainty on the kinematic distributions of the events, but not overall normalisations. They are measured for the shape of the large- $R$  jet mass distribution, since this is the variable used in the fit to data.

In this section, truth tagging, the method used to ensure sufficient numbers of jets are available to calculate uncertainties, is described in Section 7.3.1. The different sources of background modelling uncertainty which have been assessed are described in Section 7.3.2, and details of their implementations are found in Section 7.3.3. Full descriptions of the modelling uncertainties applied for the  $V+jets$  and diboson backgrounds are given in Section 7.3.4 and Section 7.3.5, respectively. Finally,

brief descriptions of the modelling uncertainties applied for the  $t\bar{t}$  and single-top backgrounds are given in Section 7.3.6, while the signal modelling uncertainties are outlined in Section 7.3.7.

### 7.3.1 Truth Tagging

Modelling studies involving  $c$ - and light-jets is hampered by the low number of events available after the analysis selection is applied, due to the high rejection rates of the  $b$ -tagging algorithm MV2c10. For modelling studies, truth tagging (TT) is therefore employed to ensure sufficient numbers of jets are available to calculate uncertainties. TT works by computing a 2-dimensional efficiency map using the jet  $p_T$  and  $\eta$ . The two leading track-jets associated to the large- $R$  jet are weighted based on their probability to pass the  $b$ -tagging selection, rather than being required to explicitly pass the  $b$ -tagging requirement.

### 7.3.2 Sources of Systematic Modelling Uncertainties

This section briefly describes the different sources of uncertainty in the analysis, and how each is implemented.

#### QCD Scales

The  $V + \text{jets}$  matrix element calculations contains infrared and ultraviolet divergences. These are handled by introducing arbitrary parameters corresponding to the renormalisation scale ( $\mu_R$ ) and factorisation scale ( $\mu_F$ ). Physical observables are not dependent on these parameters when using the infinite perturbation series expansion, however at some fixed order in QCD a limited dependence is present. To assess the impact of this, both  $\mu_R$  and  $\mu_F$  are independently varied from their nominal values by factors of 0.5 and 2 to account for higher order corrections to the calculation of the matrix element used to simulate the process.

## PDF Sets

Parton distribution functions (PDFs) specify the probability of finding a parton with a given momentum inside a hadron (in this case, colliding protons). PDFs have to be derived from data and are a significant source of uncertainty in analyses of hadronic collisions. There are three components to the PDF uncertainties: the statistical and systematic errors on the underlying data used to derive the PDFs, the theory which is used to describe them (which is based on some fixed order perturbative QCD expansion), and finally the procedure which is used to extract the PDFs from the data. PDF-related uncertainties were derived following Ref. [181]. This involves considering 100 PDF replicas which, when combined, form a central value and associated uncertainty, and also in parallel direct changes to the central values of PDFs using the MMHT2014 [187] and CT14NLO [188] PDF sets.

## Event Generator

The choice of parton shower (PS) and underlying event (UE) generators can affect the analysis outcome. Changing these models modifies several aspects of the event generation at the same time, such as the accuracy of matrix element predictions and different approaches to parton showering. This change tends to lead to the largest discrepancy with respect to the nominal samples.

## Resummation and Merging Scales

Resummation is a technique used in QCD to help cope with calculations involving disparate energy scales, and involves the introduction of an associated resummation scale, the choice of which introduces some systematic uncertainty into the model. Parton showering models are accurate when simulating low- $p_T$  radiation, however inaccuracies start to arrive when simulating hard emissions. To combat this, parton showering models utilise more precise matrix element calculations above some momentum threshold. The choice of threshold, or *merging scale* introduces some uncertainty into the final result. Resummation (QSF) and merging (CKKW) scale variations are available for a subset of the SHERPA samples. The number of available events is significantly lower than the number of events in the nominal sample, and no

statistically significant discrepancy with respect to the nominal samples is observed. The corresponding uncertainties are therefore neglected.

### 7.3.3 Implementation of Variations

Modelling variations are implemented in various ways, depending on the associated uncertainty. Table 7.5 lists the different sources of uncertainty described in Section 7.3.2 and for each lists the implementation.

Source of Uncertainty	Implementation
Renormalisation scale ( $\mu_R$ )	Internal weights
Factorisation scale ( $\mu_F$ )	Internal weights
PDF set	Internal weights
Parton Shower (PS) models	Alternative samples
Underlying Event (UE) models	Alternative samples
Resummation scale (QSF)	Parameterisation
Merging scale (CKKW)	Parameterisation

**Table 7.5:** Different sources of uncertainty (i.e. variations in the model) considered for the  $V+jets$  background, and the corresponding implementation. For each uncertainty, acceptance and shape uncertainties are derived.

As production of large numbers of MC events is computationally expensive, a technique in state of the art simulation packages is to store some sources of variation as internal weights, which can be generated alongside the nominal samples, saving computation time. The nominal sample then effectively contains information about an ensemble of different samples, corresponding to different model parameters, which are accessible via reweightings.

While the inclusion of internal weight variation in some MC event generators has decreased simulation times and increased available statistics, there are currently some sources of systematic uncertainty that are unable to be stored as internal weight variations due to technical limitations in some of these generators. Two examples are the choice of resummation and merging scales. A method to parameterise the systematic variation using one sample, and to then apply this parameterisation to another sample, has been developed by ATLAS [189]. This method was used to

derive resummation and merging uncertainties for the nominal SHERPA 2.2.1 sample, using a previous SHERPA 2.1 alternate sample. The resulting uncertainties were studied and found to be negligible in comparison with systematics from other sources.

### 7.3.4 Vector Boson + Jets Modelling

After event selection, the  $V+$ jets background is a dominant background in all three analysis channels as described in Section 7.1.4. The  $V+$ jets samples are split into categories depending on the truth flavour of the track-jets which are ghost-associated to the large- $R$  jet Higgs candidate. The categories are  $V+bb$ ,  $V+bc$ ,  $V+bl$ ,  $V+cc$ ,  $V+cl$ ,  $V+ll$ , and  $V+hf$  refers collectively to the  $bb$ ,  $bc$ ,  $bl$ , and  $cc$  categories.  $V+bb$  is dominant generally accounting for 80% of the events, while the combined  $V+hf$  accounts for around 90% of the events. The full flavour composition breakdown for each channel and analysis region are given in Tables 7.6, 7.8 and 7.9.

The nominal MC event generator used for  $V+$ jets events is SHERPA 2.2.1, while MADGRAPH5\_AMC@NLO+PYTHIA8 (which uses a different parton showering model) is used as an alternative generator.

Sample	$Mp_T^V$ HP SR	$Hp_T^V$ HP SR	$Mp_T^V$ LP SR	$Hp_T^V$ LP SR	$Mp_T^V$ CR	$Hp_T^V$ CR
<b>Wbb</b>	$79.3\% \pm 4.4\%$	$75.0\% \pm 8.6\%$	$77.4\% \pm 2.5\%$	$71.7\% \pm 4.5\%$	$68.0\% \pm 7.6\%$	$63.5\% \pm 14.0\%$
<b>Wbc</b>	$6.6\% \pm 0.9\%$	$3.4\% \pm 1.7\%$	$6.2\% \pm 0.5\%$	$5.3\% \pm 0.9\%$	$14.5\% \pm 3.2\%$	$3.4\% \pm 3.2\%$
<b>Wbl</b>	$3.9\% \pm 0.9\%$	$11.4\% \pm 3.5\%$	$4.5\% \pm 0.5\%$	$8.7\% \pm 1.4\%$	$9.8\% \pm 2.2\%$	$9.1\% \pm 3.8\%$
<b>Wcc</b>	$5.1\% \pm 1.7\%$	$6.8\% \pm 2.4\%$	$7.1\% \pm 1.0\%$	$6.3\% \pm 1.4\%$	$4.2\% \pm 2.4\%$	$12.3\% \pm 7.0\%$
<b>Wcl</b>	$2.3\% \pm 1.4\%$	$2.4\% \pm 2.1\%$	$3.4\% \pm 0.7\%$	$5.2\% \pm 1.5\%$	$2.6\% \pm 1.5\%$	$3.4\% \pm 2.1\%$
<b>Wl</b>	$2.9\% \pm 1.0\%$	$0.9\% \pm 1.6\%$	$1.3\% \pm 0.7\%$	$2.8\% \pm 0.7\%$	$0.9\% \pm 0.6\%$	$8.4\% \pm 5.1\%$
Events	<b><math>187.5 \pm 7.7</math></b>	<b><math>38.2 \pm 3.1</math></b>	<b><math>429.5 \pm 10.0</math></b>	<b><math>97.8 \pm 4.2</math></b>	<b><math>33.8 \pm 2.5</math></b>	<b><math>8.3 \pm 1.2</math></b>

**Table 7.6:** 0-lepton  $W+$ jets nominal sample flavour composition and total event yield [190].  $Mp_T^V$  refers to the medium  $p_T^V$  region, and  $Hp_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

### $V+$ jets Acceptance Uncertainties

Several different types of acceptance uncertainties have been calculated and implemented as nuisance parameters in the fit. These account for the uncertainty

Sample	$Mp_T^V$ HP SR	$Hp_T^V$ HP SR	$Mp_T^V$ LP SR	$Hp_T^V$ LP SR	$Mp_T^V$ CR	$Hp_T^V$ CR
<b>Wbb</b>	$77.2\% \pm 2.6\%$	$72.4\% \pm 4.3\%$	$77.8\% \pm 1.8\%$	$69.3\% \pm 2.5\%$	$64.6\% \pm 4.9\%$	$53.5\% \pm 6.3\%$
<b>Wbc</b>	$7.4\% \pm 0.7\%$	$7.3\% \pm 1.1\%$	$6.6\% \pm 0.4\%$	$6.3\% \pm 0.6\%$	$13.7\% \pm 1.9\%$	$16.4\% \pm 3.7\%$
<b>Wbl</b>	$4.0\% \pm 0.5\%$	$6.7\% \pm 1.1\%$	$5.1\% \pm 0.3\%$	$8.7\% \pm 0.8\%$	$10.3\% \pm 1.7\%$	$14.6\% \pm 3.0\%$
<b>Wcc</b>	$6.2\% \pm 1.1\%$	$5.5\% \pm 1.7\%$	$6.6\% \pm 0.6\%$	$6.4\% \pm 0.7\%$	$4.5\% \pm 1.7\%$	$9.5\% \pm 3.0\%$
<b>Wcl</b>	$3.6\% \pm 0.8\%$	$4.2\% \pm 1.8\%$	$2.8\% \pm 0.5\%$	$6.2\% \pm 0.8\%$	$4.6\% \pm 1.2\%$	$4.4\% \pm 1.5\%$
<b>Wl</b>	$1.5\% \pm 0.5\%$	$3.9\% \pm 1.3\%$	$1.1\% \pm 0.2\%$	$3.1\% \pm 0.5\%$	$2.3\% \pm 1.2\%$	$1.6\% \pm 0.6\%$
Events	<b><math>477.1 \pm 11.7</math></b>	<b><math>147.5 \pm 6.4</math></b>	<b><math>784.7 \pm 12.3</math></b>	<b><math>301.8 \pm 7.2</math></b>	<b><math>68.7 \pm 3.5</math></b>	<b><math>26.9 \pm 2.0</math></b>

**Table 7.7:** 1-lepton  $W+jets$  nominal sample flavour composition and total event yield [190].  $Mp_T^V$  refers to the medium  $p_T^V$  region, and  $Hp_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

Channel	$Mp_T^V$ HP SR	$Hp_T^V$ HP SR	$Mp_T^V$ LP SR	$Hp_T^V$ LP SR	$Mp_T^V$ CR	$Hp_T^V$ CR
<b>Zbb</b>	84.56%	81.84%	82.37%	76.06%	66.12%	63.18%
<b>Zbc</b>	6.03%	6.98%	5.80%	7.46%	15.04%	14.30%
<b>Zbl</b>	4.06%	6.55%	3.83%	6.59%	12.66%	12.81%
<b>Zcc</b>	3.68%	3.40%	5.82%	3.75%	3.36%	3.38%
<b>Zcl</b>	1.23%	0.44%	1.47%	3.97%	1.82%	4.95%
<b>Zl</b>	0.44%	0.78%	0.70%	2.16%	1.00%	1.38%
Events	$259.91 \pm 4.86$	$66.12 \pm 2.04$	$420.45 \pm 5.73$	$141.97 \pm 2.50$	$43.49 \pm 1.73$	$16.07 \pm 0.83$

**Table 7.8:** 0-lepton  $Z+jets$  nominal sample flavour composition and total event yield [190].  $Mp_T^V$  refers to the medium  $p_T^V$  region, and  $Hp_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

Channel	$Mp_T^V$	$Hp_T^V$	$p_T^V$ inclusive
<b>Zbb</b>	80.80%	76.95%	79.76%
<b>Zbc</b>	8.10%	6.26%	7.60%
<b>Zbl</b>	4.95%	7.06%	5.52%
<b>Zcc</b>	3.97%	4.46%	4.10%
<b>Zcl</b>	1.61%	3.60%	2.14%
<b>Zll</b>	0.57%	1.68%	0.87%
Events	$115.49 \pm 2.42$	$42.42 \pm 1.27$	$157.92 \pm 2.73$

**Table 7.9:** 2-lepton  $Z+jets$  nominal sample flavour composition and total event yield [190].  $Mp_T^V$  refers to the medium  $p_T^V$  region, and  $Hp_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

in the overall number of events in each channel, and for the migration of events between different analysis regions. The acceptance uncertainties relevant to the  $V+jets$  processes are summarised below.

- **Overall normalisation:** only relevant where normalisation cannot be left unconstrained (or “floating”, i.e. determined as part of the fit). The  $V+hf$  component is left floating in the fit, with independent normalisations used for  $W+hf$  and  $Z+hf$ . The normalisations are mainly determined by the 1-lepton (for  $W+hf$ ) and 2-lepton (for  $Z+hf$ ) regions respectively and then extrapolated to the 0-lepton channel. The negligible  $V+jets$  backgrounds were constrained to their cross-sections in the fit.
- **SR-to-CR relative acceptance:** the uncertainty on the relative number of  $V+jets$  events in the signal and control regions.
- **HP-to-LP relative acceptance:** the uncertainty on the relative number of  $V+jets$  events in the HP and LP SRs.
- **Medium-to-high  $p_T^V$  relative acceptance:** the uncertainty on the relative number of  $V+jets$  events in the medium and high  $p_T^V$  bins.
- **Flavour relative acceptance:** for each flavour  $V+xx$ , where  $xx \in \{bc, bl, cc\}$  the ratio of  $V+xx/V+bb$  events is calculated. This corresponds to the uncertainty on the heavy flavour composition of the  $V+hf$  background.
- **Channel relative acceptance:** the uncertainty on the relative number of  $V+jets$  events between the channels.

The uncertainties arising from the different sources described in Section 7.3.2 are summed in quadrature to give a total uncertainty on each region. A summary of the different acceptance uncertainties that were derived and subsequently applied in the fit are given in Table 7.10. An effort has been made, wherever possible, to harmonise similar uncertainties across different analysis regions and channels.

## $V+jets$ Shape Uncertainties

In order to derive shape uncertainties for a given background or signal process, normalised distributions of the reconstructed large- $R$  Higgs candidate jet mass  $m_J$

V+jets Acceptance Uncertainties				
Boson	W		Z	
Channel	0L	1L	0L	2L
SR-to-CR	90% <sup>†</sup>	40% <sup>†</sup>	40%	-
HP-to-LP SR		18%	18%	-
Medium-to-high $p_T^V$	30%	10%*	10%	
Channel relative acceptance	20%	-	16%	-
Vbc/Vbb	30%			
Vbl/Vbb	30%			
Vcc/Vbb	20%			
Vcl Norm.	30%			
VI Norm.	30%			

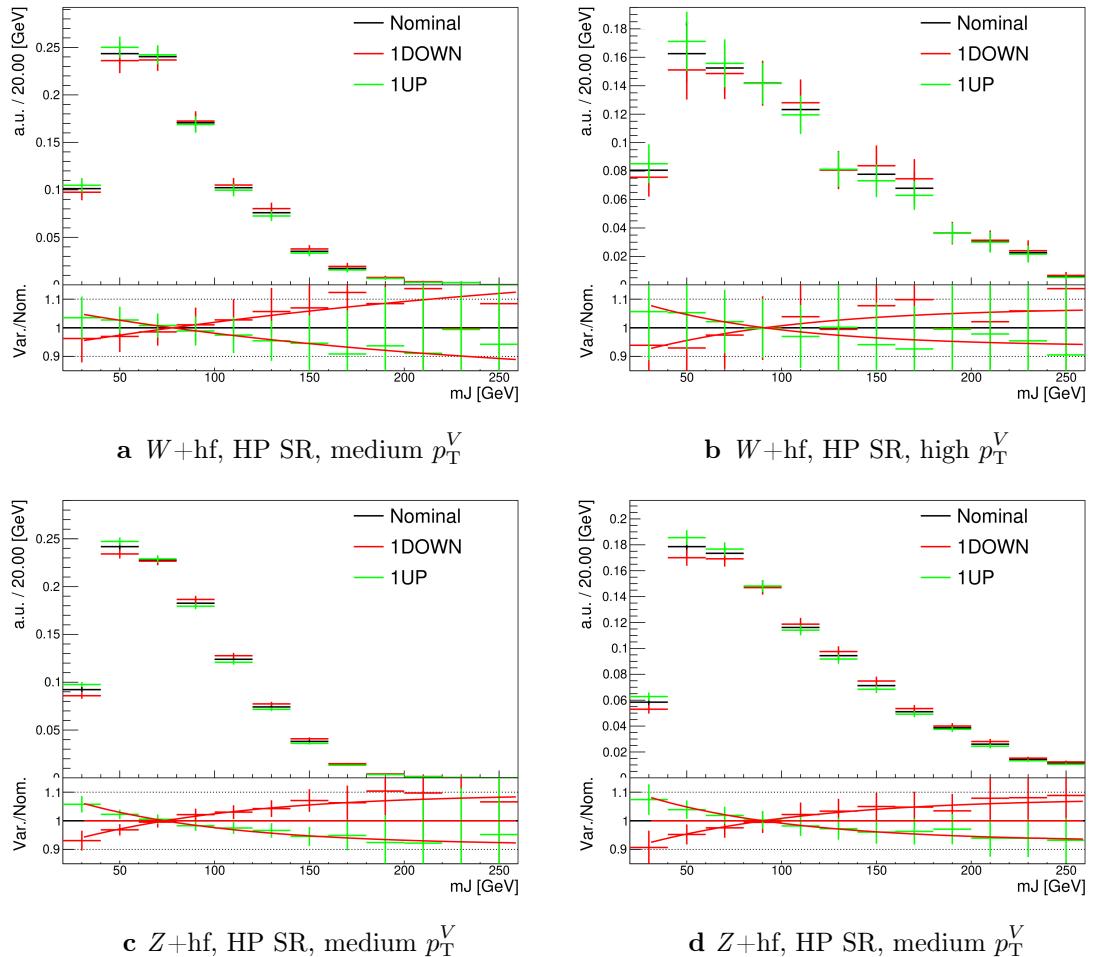
**Table 7.10:**  $V+jets$  acceptance uncertainties.  $W+jets$  SR and CR uncertainties marked with a superscript  $\dagger$  are correlated. The 1L  $W+jets$  medium-to-high  $p_T^V$  uncertainty marked by \* is applied as independent and uncorrelated NPs in both HP and LP signal regions [190].

are compared for the nominal sample and variations. For each variation, the ratio of the variation to nominal is calculated, the up and down variations (if present) are symmetrised, and an analytic function is used to parameterise the ratio. If different analysis regions or channels show the same pattern of variation, a common uncertainty is assigned.

An example of a significant source of uncertainty, arising from the choice of factorisation scale  $\mu_R$ , is shown in Fig. 7.2. The HP SRs in the medium and high  $p_T^V$  bins are shown for the 0-lepton channel for the  $W+hf$  and  $Z+hf$  jet backgrounds. The 0- and 1-lepton channels for the  $W+hf$  contribution and the 0- and 2-lepton channels for the  $Z+hf$  contribution were found to have compatible shapes in  $m_J$ , and so were jointly measured. An exponential function  $e^{p_0 + p_1 x} + p_2$  has been fitted to the ratio of the normalised distributions. The magnitude of the variation is  $p_T^V$  dependent, and so separate uncertainties are implemented in the fit for each  $p_T^V$  region.

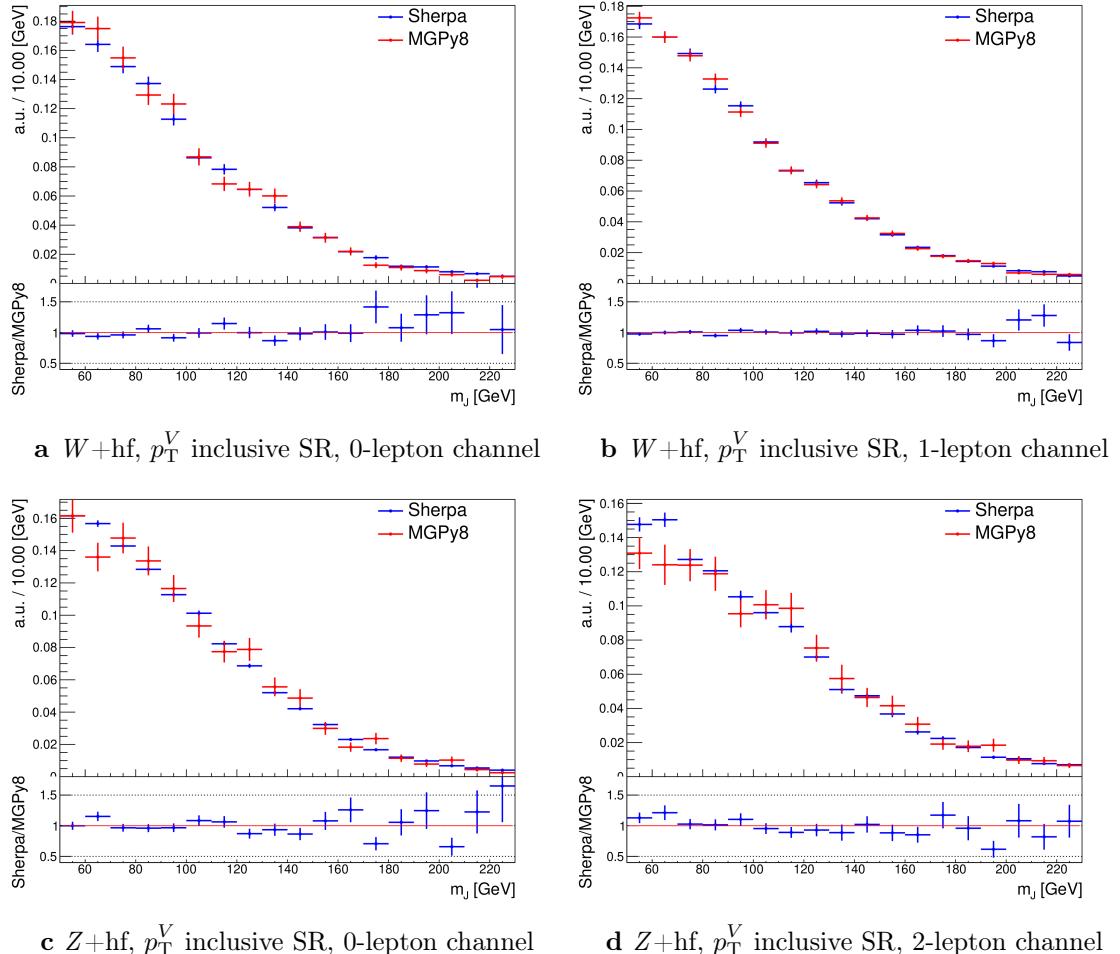
The shape uncertainties for  $\mu_R$  were derived on the SRs but are also applied to the CRs, as the low statistics in the CRs make it difficult to derive dedicated shape uncertainties. All the shape uncertainties are fully correlated across regions.

As an example where no significant deviation is seen, a comparison of the  $m_J$  shapes between SHERPA and MADGRAPH is shown in Fig. 7.3. The plots are split by process



**Figure 7.2:** Leading large- $R$  jet mass for the  $Z$  and  $W + hf$  processes in the HP SR of the 0-lepton channel [190]. The renormalisation scale  $\mu_r$  has been varied by a factor of 0.5 (1DOWN) and 2 (1UP). An exponential function is fitted to the ratio between the nominal and alternate samples.

and channel, but merged in SR and  $p_T^V$  bins reflecting similarities between the  $m_J$  shapes and variations. Due to the lack of statistically significant variation between the samples, no additional shape uncertainty was added to the fit in this case.



**Figure 7.3:** Leading large- $R$  jet  $m_J$  inclusive in  $p_T^V$  for the  $V + \text{hf}$  process modelled using both the SHERPA + PYTHIA8 (blue) and MADGRAPH + PYTHIA8 (red) samples [190].

The impacts of variations in the factorisation scale  $\mu_F$  and the choice of PDF set on  $m_J$  shape were also found to be negligible in comparison with  $\mu_R$  and are hence no additional uncertainty was added to the fit.

### 7.3.5 Diboson Modelling

The procedure to derive the uncertainties for the diboson background generally follows that of  $V+jets$ . The nominal diboson samples are generated using SHERPA 2.2.1 (except for  $gg \rightarrow VV$  which uses SHERPA 2.2.2) with the NNPDF3.0NNLO tune. Alternate samples were generated using POWHEG interfaced with PYTHIA8, using the AZNLO shower tune with the CTEQ6L1 PDFs [191]. Unlike SHERPA, POWHEG models the off-shell  $Z$  contribution at NLO.

Acceptance and shape uncertainties are derived in an analogous fashion to  $V+jets$  as described below.

#### Diboson Acceptance Uncertainties

Diboson acceptance uncertainties are summarised in Table 7.11. Variations from  $\mu_R$ ,  $\mu_F$ , PDF choice and an alternative generator are considered and are combined via a sum in quadrature as described in Section 7.3.4. The largest modification to the nominal acceptance is from the POWHEG+PYTHIA8 alternate sample. Since the diboson contribution to the  $t\bar{t}$  control region is negligible, no SR-to-CR relative acceptance uncertainty is necessary.

For the  $WZ$  contribution, uncertainties are derived using the 1-lepton channel and applied to all three channels. The 1-lepton channel was used as it has the largest amount of available statistics, with the compatibility checked between the derived uncertainties and the other channels. An additional 8% channel migration uncertainty is applied on the  $WZ$  0-lepton channel. For the  $ZZ$  contribution, the normalisation uncertainty is calculated using the 2-lepton channel and applied to all three channels. 30% and 18% channel migration uncertainties are applied in the 0- and 1-lepton channels respectively. The 0- and 1-lepton channels were found to have a similar HP-to-LP relative acceptance uncertainty of 18%. The 1-lepton medium-to-high  $p_T^V$  relative acceptance is estimated using the 2-lepton channel, since the 1-lepton channel had an insufficient number of events to estimate the uncertainty directly.

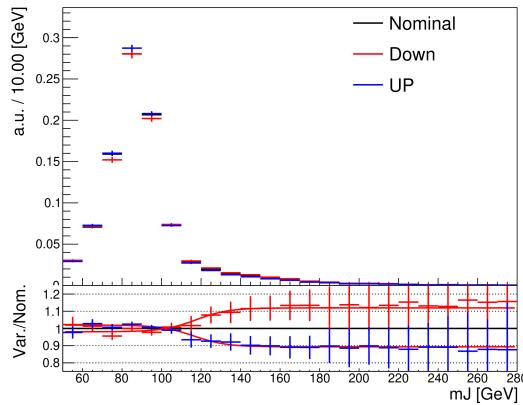
Since the contribution from  $WW$  is negligible, dedicated studies are not performed, but a 25% normalisation uncertainty is applied in all the three channels which is based on the modelling studies performed for the previous analysis [89].

Diboson Acceptance Uncertainties						
Bosons	WZ			ZZ		
Channel	0L	1L	2L	0L	1L	2L
Normalisation	16%			10%		
HP-to-LP SR	18%			18%		
Medium-to-high $p_T^V$	10%			6%	18%	
Channel Relative acceptance	8%	-	-	30%	18%	-

**Table 7.11:** Diboson acceptance uncertainties. All uncertainties except channel extrapolation uncertainties are fully correlated between the  $ZZ$  and  $WZ$  processes and channels [190].

### Diboson Shape Uncertainties

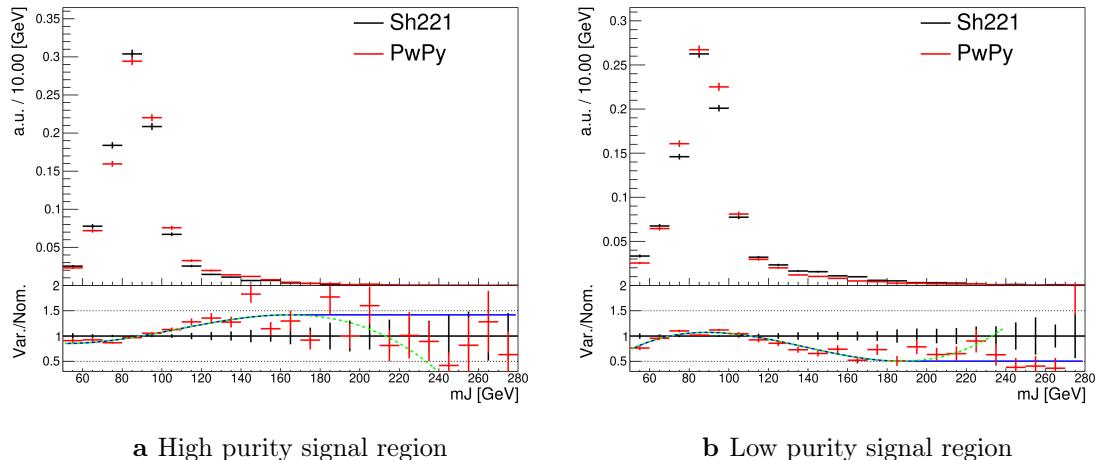
Diboson shape uncertainties are derived in a similar fashion to  $V+jets$ . Only the uncertainties associated with the systematic variations of  $\mu_R$  and the alternate event generator have a non-negligible impact on the  $m_J$  shape. Variation of  $\mu_R$  produces consistent  $m_J$  shape changes across all regions and channels, and hence only a single associated uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted to the symmetrised ratio.



**Figure 7.4:** Leading large- $R$  jet mass distribution for the combined  $WZ$  and  $ZZ$  processes, inclusive across all signal regions and lepton channels [190]. The renormalisation scale  $\mu_R$  has been varied by a factor of 2 (UP) and 0.5 (DOWN). The red and blue curves show the fitted results of the hyperbolic tangent function.

In the 2-lepton channel, no significant shape differences between the nominal and alternate was observed. The comparison between the nominal SHERPA and alternate

POWHEG+PYTHIA8 samples is shown in Fig. 7.5 for the 0- and 1-lepton channels for both the  $WZ$  and  $ZZ$  processes. For these channels, the shape of  $m_J$  varies in opposite directions in the LP and HP signal regions. Shapes are similar between  $p_T^V$  bins, the 0- and 1-lepton channels and for  $WZ$  and  $ZZ$ . In order to reduce the effects of statistical fluctuations on the fitted shape, these regions are merged before deriving the shape uncertainty. A third order polynomial is fitted to the ratio, and this function transitions to a constant piecewise function in the high mass region to accurately represent the shape. Dependence on the event generator was found to be negligible within statistical uncertainty in the 2-lepton channel, and so no uncertainty was applied. All diboson shape uncertainties are fully correlated in the fit.



**Figure 7.5:** Comparison of the shapes of the large- $R$  jet mass  $m_J$  between SHERPA (Sh221) (black) and POWHEG+PYTHIA8 (PwPy) (red) samples from the  $WZ$  and  $ZZ$  processes in high and low purity signal regions, integrated over  $p_T^V$  regions and the 0- and 1-lepton channels [190]. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after protection is added in the high mass region.

### 7.3.6 $t\bar{t}$ and single-top Modelling

The main features of the systematic uncertainties on the remaining two modelled backgrounds,  $t\bar{t}$  and single-top, are described below.

The modelling of the  $t\bar{t}$  background uses a POWHEG+PYTHIA8 nominal sample. Two alternate samples were considered: POWHEG+HERWIG7 (providing an alternate

parton shower model) and MADGRAPH5+PYTHIA8.2 (providing an alternate hard scatter model). Effects of initial and final state radiation (ISR and FSR, respectively) were assessed using internal weight variations in the nominal sample. Acceptance and shape uncertainties were derived for each of the variations. The largest contribution for the acceptance uncertainties is due to the matrix element calculation, with the parton showering model being second. The ISR and FSR acceptance uncertainties were found to be subdominant. Acceptance uncertainties for the single-top background are enumerated in Table 7.12. For the shape uncertainties, only the ISR and parton showering variations have non-negligible impacts on the  $m_J$  shape.

$t\bar{t}$ Acceptance Uncertainties	
HP-to-LP SR	18%
Medium-to-high $p_T^V$	20%
SR-to-CR	6%

**Table 7.12:** Acceptance uncertainties for the  $t\bar{t}$  background. The uncertainties on the 0- and 1-lepton channels are found to be similar, but are conservatively taken to be uncorrelated. The HP-to-LP SR uncertainty is applied in the HP SR. The Medium-to-high  $p_T^V$  uncertainty is applied in the high  $p_T^V$  bin, and is larger than the computed uncertainty in order to account for the known mismodelling of the  $t\bar{t}$   $p_T$  spectrum. [190].

The dominant process contributing to the single-top background is  $Wt$  production for the 0- and 1-lepton channels. The same nominal and alternate generators are used as for the  $t\bar{t}$  background. Again, ISR and FSR variations are obtained from internal weight variations in the nominal sample. At higher orders in QCD, diagrams contributing to the  $Wt$  production process can also be found in leading-order  $t\bar{t}$  production processes. To account for the arising interference effects, the diagram removal (DR) scheme in Ref. [192] was employed for the nominal sample. The uncertainty on the DR scheme was assessed using an alternate sample using a diagram subtraction (DS) method which removes interference at the generator level. The largest sources of acceptance and shape uncertainties were due to this DS-DR variation. Acceptance uncertainties for the single-top background are enumerated in Table 7.13.

Single-Top Acceptance Uncertainties	
Normalisation	20%
HP-to-LP SR	25%
Medium-to-high $p_T^V$	20%
SR-to-CR	30%
Channel Relative acceptance	20%

**Table 7.13:** Single-top acceptance uncertainties. Uncertainties are derived in the 1-lepton channel but applied in a correlated fashion to both the 0- and 1-lepton channels. The HP-to-LP SR uncertainty is applied in the HP SR. The Medium-to-high  $p_T^V$  uncertainty is applied in the high  $p_T^V$  bin. The channel relative acceptance uncertainty is applied in the 0-lepton channel. [190].

### 7.3.7 Signal Modelling

The modelling of the systematic uncertainties affecting the signal processes follows the procedure described in Refs. [28, 193–196]. The  $qq \rightarrow VH$  signal samples are generated with POWHEG BOX v2 + GoSAM at next-to-leading order (NLO) accuracy in QCD. An additional  $gg \rightarrow ZH$  sample is generated using POWHEG BOX v2 at leading order (LO) in QCD. In both cases, the generated events are interfaced with PYTHIA 8 for the parton showering modelling. An alternate HERWIG7 sample is used to assess the uncertainty on the parton showering model. Recommended systematic uncertainties on the signal production cross-sections and  $H \rightarrow b\bar{b}$  branching ratio from the LHC Higgs Cross Section Working Group are applied [197, 198]. This includes acceptance and shape uncertainties arising from missing higher-order QCD and electroweak corrections, PDF uncertainties, renormalisation and factorisation scales, and an alternate parton showering model.

## 7.4 Statistical Treatment

A binned global maximum-profile-likelihood fit of the  $m_J$  distribution is performed to extract information on the signal, combining all the analysis regions defined in Table 7.1. The signal strength  $\mu = \sigma/\sigma_{\text{SM}}$  is defined as the ratio between the observed and predicted cross-sections, where  $\mu = 0$  corresponds to the background-only hypothesis and  $\mu = 1$  corresponds to the SM prediction. This is a parameter of

interest (POI) which acts to scale the total number of signal events, and is determined during the fit procedure.

The present analysis makes use of two POIs. The first,  $\mu_{VH}^{bb}$ , is the signal strength for the  $VH$ ,  $H \rightarrow b\bar{b}$  process, the primary process under investigation. The diboson production strength  $\mu_{VZ}^{bb}$  for the  $VZ$ ,  $Z \rightarrow b\bar{b}$  process is measured simultaneously and provides a validation of the analysis apparatus used for the primary  $H \rightarrow b\bar{b}$  measurement. Alongside the two POIs, the predictive model depends on several parameters which are not the primary target of measurement, and represent the systematic uncertainties discussed previously. These parameters are called nuisance parameters (NPs), collectively referred to as  $\theta$ . Freely floating background normalisations are implemented as NPs and are also extracted during the fitting processes.

#### 7.4.1 The Likelihood Function

The statistical setup treats each bin as a Poisson counting experiment and is based on the ROOSTATS framework [199, 200]. The combined likelihood over  $N$  bins is constructed as the product of Poisson probabilities in each bin. Considering the simplified case of a single signal strength parameter  $\mu$ , and neglecting sources of systematic or statistical uncertainty, this is given by

$$\mathcal{L}(\mu) = \prod_{i=1}^N \frac{(\mu s_i + b_i)^{n_i}}{n_i!} \exp [-(\mu s_i + b_i)], \quad (7.1)$$

where  $s_i$  ( $b_i$ ) is the expected number of signal (background) events in bin  $i$ , and  $n_i$  is the number of observed data events in bin  $i$ .

#### Treatment of Uncertainties

Systematic uncertainties can modify the predicted signal and background yields  $s_i$  and  $b_i$ . Each source of systematic uncertainty is taken into account by adding an additional NP  $\theta_j$  to the likelihood in the form of a Gaussian cost function. The

combined effect of the NPs is then

$$\mathcal{L}(\theta) = \prod_{j=1}^{N_\theta} \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left[ \frac{-(\theta_j - \hat{\theta}_j)^2}{2\sigma_j^2} \right], \quad (7.2)$$

where  $N_\theta$  is the number of NPs,  $\theta_j$  is the nominal value of the  $j$ th NP,  $\hat{\theta}_j$  is the fitted value, and  $\sigma_j$  is the corresponding associated prior uncertainty on the NP. As the fitted value  $\hat{\theta}_j$  deviates from its nominal value, a cost is introduced. The presence of NPs modifies the likelihood as

$$\mathcal{L}(\mu) \rightarrow \mathcal{L}(\mu, \theta) = \mathcal{L}(\mu)\mathcal{L}(\theta). \quad (7.3)$$

The predicted signal and background yields are also modified by the presence of the NPs with

$$s_i \rightarrow s_i(\theta), \quad b_i \rightarrow b_i(\theta). \quad (7.4)$$

For NPs which are left freely floating in the fit, no corresponding Gaussian constraints are added to the likelihood.

The pull of a NP is defined as the difference between the fitted value  $\hat{\theta}_j$  and the nominal value  $\theta_j$ , divided by the uncertainty on the NP  $\sigma_j$ . To obtain the uncertainty on the pull of a NP, the following procedure is used. The Hessian matrix  $\mathbf{H}$  is calculated as

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \theta_1^2} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_n} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \mathcal{L}}{\partial \theta_n \partial \theta_1} & \frac{\partial^2 \mathcal{L}}{\partial \theta_n \partial \theta_2} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \theta_n^2} \end{bmatrix}. \quad (7.5)$$

Taking the inverse of the Hessian matrix  $\mathbf{H}^{-1}$  yields the covariance matrix, from which the post-fit uncertainties on the different NPs can be extracted. If the post-fit

uncertainty is smaller than the nominal uncertainty, additional information about the NP has been extracted by the fit, and NP is said to be *constrained*.

Following Ref. [201], the statistical uncertainty on the simulated events is implemented using a dedicated NP for each bin which can scale the background yield in that bin. Statistical NPs are also implemented using a Gaussian constraint.

### Smoothing and Pruning

To simplify the fit to reduce and improve its robustness, systematic uncertainties are smoothed and pruned. Smoothing accounts for the large statistical uncertainty present in some samples that can lead to unphysical fluctuations in the shape systematics. The smoothing procedure relies on the assumption that the impact of systematics should be approximately monotonic and correlated between neighbouring bins.

In addition to smoothing, pruning is the process of removing from the fit those systematics which only have a very small effect. This improves the stability of the fit by reducing the number of degrees of freedom. Acceptance uncertainties are pruned in a given region if they have a variation of less than 0.5%, or if the up and down variations have the same sign in that region. Shape uncertainties are pruned in a given region if the deviation in each bin is less than 0.5% in that region. In addition, acceptance and shape uncertainties are neglected in a given region for any background which makes up less than 2% of the total background in that region.

### Fit Procedure and Statistical Tests

The best-fit value of  $\mu$ , denoted  $\hat{\mu}$ , is obtained via an unconditional maximisation of the likelihood. The likelihood is also used to construct a statistical test which can confirm or reject the background-only hypothesis. The test statistic  $q_\mu$  is constructed from the profile likelihood ratio,

$$q_\mu = -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad (7.6)$$

where  $\hat{\mu}$  and  $\hat{\theta}$  are chosen to maximise the likelihood  $\mathcal{L}$ , and the profile value  $\hat{\theta}_\mu$  is obtained from a conditional maximisation of the likelihood for a specific choice of  $\mu = 0$  corresponding to the background-only hypothesis.

The test statistic is used to construct a  $p$ -value which is used to probe the background-only hypothesis. The  $p$ -value is typically reported in terms of the significance  $Z$ , defined as the number of standard deviations for a Gaussian Normal distribution which will produce a one-sided tail integral equal to the  $p$ -value, as in

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z). \quad (7.7)$$

Typically a value of  $Z = 3$  constitutes *evidence* of a processes, while  $Z = 5$  is required for a *discovery*, or *observation*. Alongside the  $p$ -value, the best-fit value of the signal strength  $\hat{\mu}$  and its corresponding uncertainty are quoted, and compared to their expected values. More detail on the statistical methodology can be found in Ref. [202].

### 7.4.2 Background Normalisations

The backgrounds which can be constrained by the fit are left freely floating and the corresponding normalisation factors are extracted. Normalisation factors (NF), represent the value by which the predicted normalisations are scaled, and are implemented for the dominant backgrounds ( $t\bar{t}$ ,  $Z+hf$ ,  $W+hf$ ). The NFs are also subdivided into different regions of phase-space for  $t\bar{t}$ , given it is possible to obtain a strong constraint in the individual channels. This also removes the need for an extrapolation uncertainty.

The normalisations and shapes of all other backgrounds, with the exception of the multijet background which is estimated using a data driven technique, are initialised using the nominal samples and the state-of-the art process normalisations, as outlined in Table 7.4.

### 7.4.3 Asimov Dataset & Expected Results

The Asimov dataset is constructed by replacing the data with the sum of the signal and background predictions  $n_i = s_i + b_i$ . A fit to this dataset using the nominal values of the NPs from the simulation will recover the input values and is useful for studying the expected constraints on and correlations between the NPs.

Furthermore, a conditional fit to the Asimov dataset can be performed using values of the background NPs which are determined from an unconditional fit to data. The signal NPs and POIs are fixed at their nominal values from the SM simulation. The result of this fit can be used to calculate expected (median) significances given a more realistic background model, which can be compared to their observed values, as is done in Section 7.5.2.

## 7.5 Results

In the present analysis, the two signal strength parameters  $\mu_{VH}^{bb}$  and  $\mu_{VZ}^{bb}$  are extracted from a simultaneous maximisation of the likelihood described in Section 7.4. The results of the analysis are summarised in this section. Post-fit  $m_J$  distributions are shown in Section 7.5.1. The observed signal strengths are given in Section 7.5.2, along with observed and expected significances. Finally in Section 7.5.3 the impact of systematic uncertainties on the results is examined.

### 7.5.1 Post-fit Results

In addition to the observed significance and signal strength, it is also necessary to study the post-fit  $m_J$  yields and distributions to compare the level of the agreement between the simulation (using the best-fit values of the signal strength  $\hat{\mu}$  and the NP  $\hat{\theta}$ ) and the data. The best-fit values  $\hat{\mu}$  and  $\hat{\theta}$  are obtained from an unconditional fit to data over all analysis regions. The post-fit background normalisation factors extracted from the unconditional fit to data fit are shown in Table 7.14, and the post-fit yields are presented in Table 7.15, Table 7.16, and Table 7.17 for the 0-, 1- and 2-lepton channels, respectively. The uncertainty on the overall background may

be less than the sum of the individual uncertainties due to the correlations between the different background components.

Process	Normalisation factor
$t\bar{t}$ 0-lepton	$0.88 \pm 0.10$
$t\bar{t}$ 1-lepton	$0.83 \pm 0.09$
$W+hf$	$1.12 \pm 0.14$
$Z+hf$	$1.32 \pm 0.16$

**Table 7.14:** Factors applied to the nominal normalisations of the  $t\bar{t}$ ,  $W+hf$ , and  $Z+hf$  backgrounds, as obtained from the likelihood fit [142]. The errors represent the combined statistical and systematic uncertainties.

Processes	$250 \text{ GeV} < p_T^V \leq 400 \text{ GeV}$			$p_T^V > 400 \text{ GeV}$		
	HP	LP	CR	HP	LP	CR
Signal	$21.93 \pm 11.17$	$18.99 \pm 9.76$	$1.05 \pm 0.54$	$5.69 \pm 2.88$	$5.85 \pm 3.01$	$0.33 \pm 0.17$
$W+t$	$14.70 \pm 5.37$	$45.55 \pm 19.44$	$17.18 \pm 8.09$	$2.03 \pm 0.98$	$8.93 \pm 6.33$	$3.76 \pm 2.49$
other $t+X$	$0.79 \pm 0.03$	$3.18 \pm 0.66$	$4.51 \pm 1.28$	-	$0.66 \pm 0.03$	$0.11 \pm 0.00$
$t\bar{t}$	$75.19 \pm 13.60$	$423.85 \pm 36.12$	$539.21 \pm 31.39$	$7.54 \pm 1.77$	$38.20 \pm 6.75$	$44.07 \pm 7.43$
$VZ$	$77.01 \pm 17.09$	$87.70 \pm 19.36$	$6.16 \pm 1.56$	$17.30 \pm 4.10$	$28.77 \pm 6.55$	$2.79 \pm 0.72$
$WW$	-	$2.15 \pm 0.05$	$0.24 \pm 0.01$	$0.33 \pm 0.02$	$1.80 \pm 0.06$	-
$Whf$	$100.78 \pm 20.01$	$331.31 \pm 59.54$	$29.97 \pm 21.85$	$20.19 \pm 6.24$	$59.82 \pm 17.91$	$6.61 \pm 5.09$
$Wcl$	$5.13 \pm 2.31$	$8.44 \pm 3.24$	$0.46 \pm 0.01$	$0.99 \pm 0.69$	$2.77 \pm 1.14$	$0.19 \pm 0.07$
$Wl$	$5.61 \pm 3.93$	$4.61 \pm 2.45$	$0.16 \pm 0.00$	$1.41 \pm 2.06$	$2.67 \pm 1.67$	$0.57 \pm 0.36$
$Zhf$	$318.76 \pm 35.27$	$548.71 \pm 61.84$	$76.97 \pm 21.47$	$86.79 \pm 10.63$	$184.99 \pm 21.43$	$25.76 \pm 7.43$
$Zcl$	$3.97 \pm 1.63$	$6.74 \pm 2.68$	$0.83 \pm 0.02$	-	$6.36 \pm 2.73$	$0.93 \pm 0.41$
$Zl$	$1.34 \pm 0.67$	$3.61 \pm 2.14$	$0.42 \pm 0.01$	$1.05 \pm 0.63$	$3.68 \pm 2.47$	$0.29 \pm 0.16$
Data	623	1493	683	146	330	85
Background	$603 \pm 25$	$1466 \pm 36$	$676 \pm 25$	$138 \pm 9$	$339 \pm 15$	$85 \pm 7$

**Table 7.15:** Post-fit yields in the 0-lepton channel. Combined statistical and systematic uncertainties are shown [190].

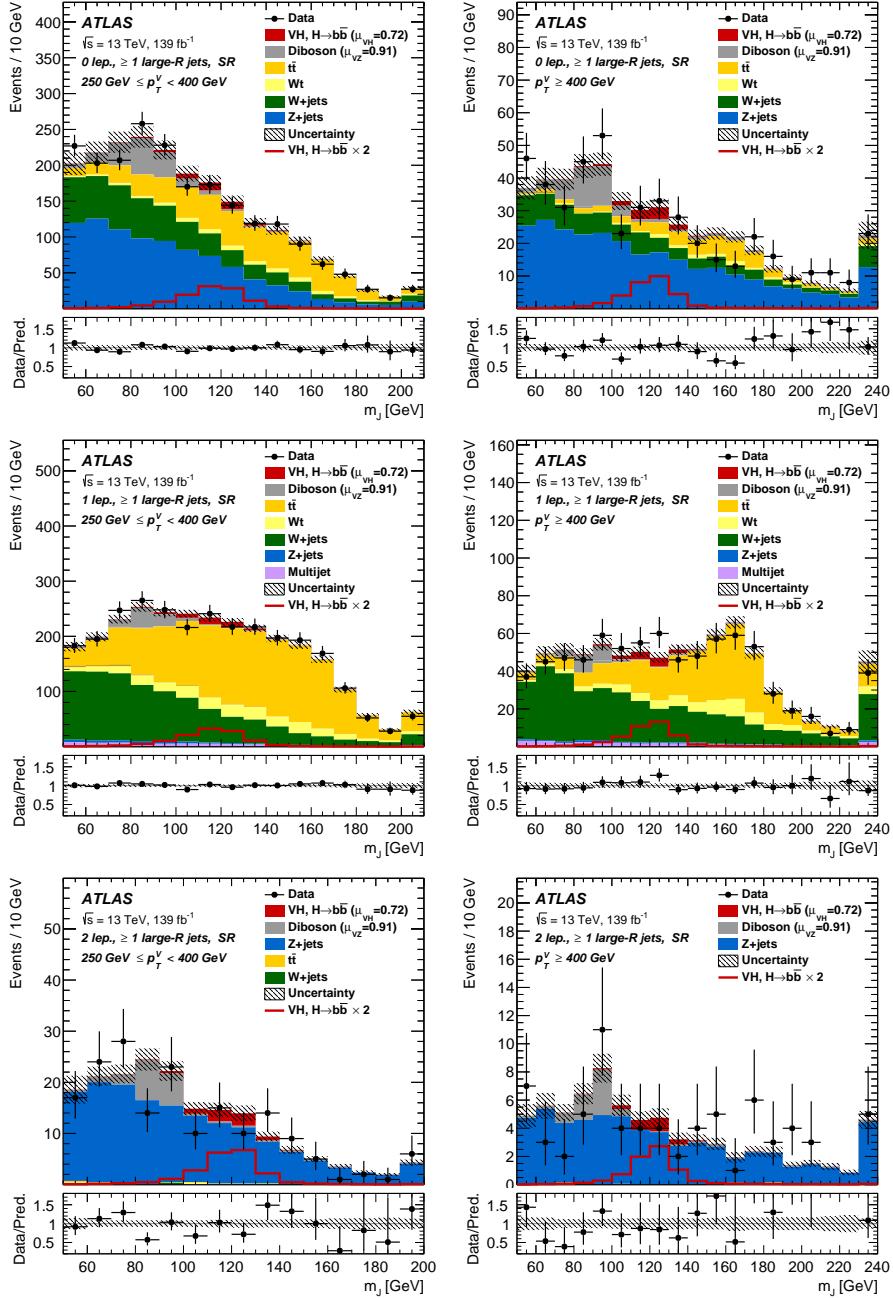
Post-fit  $m_J$  distributions are given for the signal regions in the 0-, 1- and 2-lepton channels in Fig. 7.6. The LP and HP regions are merged for the 0- and 1-lepton channels for the sake of simplicity. In general there is a good level of agreement between the simulation and data, indicating the fit model is performing as expected. Fig. 7.7 shows the post-fit plots for the  $t\bar{t}$  control regions. Again, a good level of agreement is observed given the statistical uncertainties on the distributions. The binning shown in Figs. 7.6 and 7.7 is the same as that used in the fit.

Processes	$250 \text{ GeV} < p_T^V \leq 400 \text{ GeV}$			$p_T^V > 400 \text{ GeV}$		
	HP	LP	CR	HP	LP	CR
Signal	$24.23 \pm 12.34$	$18.02 \pm 9.29$	$0.88 \pm 0.45$	$7.84 \pm 3.96$	$7.50 \pm 3.87$	$0.39 \pm 0.20$
W+t	$64.35 \pm 21.12$	$159.95 \pm 75.14$	$73.44 \pm 29.96$	$16.40 \pm 7.31$	$53.28 \pm 41.74$	$21.16 \pm 15.36$
other t+X	$1.92 \pm 0.48$	$16.33 \pm 0.31$	$21.89 \pm 6.18$	$0.13 \pm 0.01$	$1.70 \pm 0.06$	$3.95 \pm 1.40$
$t\bar{t}$	$234.76 \pm 30.21$	$1189.51 \pm 75.91$	$1758.08 \pm 57.99$	$50.87 \pm 7.34$	$226.85 \pm 23.98$	$340.61 \pm 25.32$
VZ	$35.94 \pm 8.87$	$56.30 \pm 13.98$	$4.93 \pm 1.38$	$8.63 \pm 2.30$	$20.02 \pm 5.29$	$2.61 \pm 0.84$
WW	-	$6.48 \pm 1.63$	-	-	$4.35 \pm 1.32$	$0.93 \pm 0.03$
Whf	$265.13 \pm 27.68$	$617.81 \pm 63.56$	$59.91 \pm 21.90$	$91.42 \pm 11.51$	$238.81 \pm 29.53$	$26.55 \pm 9.84$
Wcl	$7.33 \pm 2.95$	$13.81 \pm 5.65$	$2.10 \pm 0.04$	$6.23 \pm 2.49$	$10.17 \pm 4.09$	$0.63 \pm 0.02$
Wl	$2.99 \pm 1.47$	$5.66 \pm 3.39$	$0.65 \pm 0.01$	$2.21 \pm 1.35$	$7.67 \pm 4.98$	$0.31 \pm 0.01$
Zhf	$10.16 \pm 1.24$	$24.61 \pm 2.46$	$3.45 \pm 0.41$	$2.12 \pm 0.30$	$6.56 \pm 0.79$	$0.98 \pm 0.12$
Zcl	$0.02 \pm 0.00$	$0.75 \pm 0.02$	-	-	$0.33 \pm 0.01$	$0.02 \pm 0.00$
Zl	-	$0.49 \pm 0.01$	$0.03 \pm 0.00$	$0.30 \pm 0.19$	$0.23 \pm 0.01$	$0.02 \pm 0.00$
ggWW	-	$0.35 \pm 0.01$	$0.27 \pm 0.01$	$0.15 \pm 0.02$	$0.33 \pm 0.01$	-
MultiJet	$17.04 \pm 8.87$	$44.29 \pm 22.82$	$21.78 \pm 11.22$	$7.81 \pm 4.50$	$21.85 \pm 12.73$	$7.86 \pm 4.01$
Data	668	2161	1946	185	597	410
Background	$640 \pm 26$	$2136 \pm 44$	$1947 \pm 43$	$186 \pm 11$	$592 \pm 21$	$406 \pm 18$

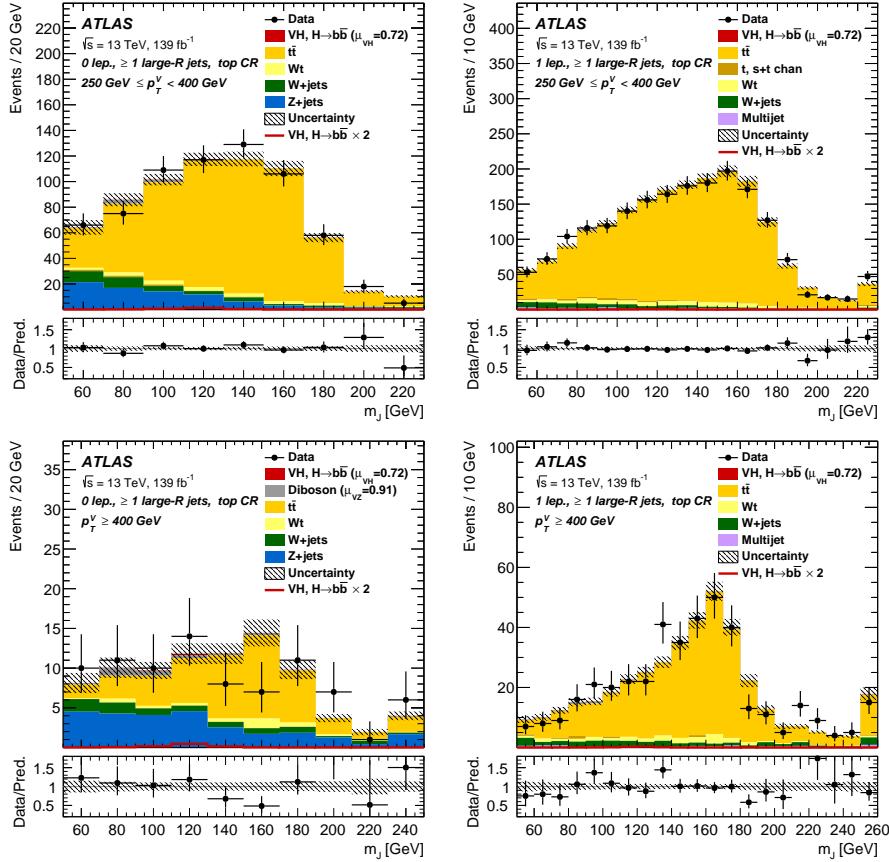
**Table 7.16:** Post-fit yields in the 1-lepton channel. Combined statistical and systematic uncertainties are shown [190].

Processes	$250 \text{ GeV} < p_T^V \leq 400 \text{ GeV}$	$p_T^V > 400 \text{ GeV}$
	SR	SR
Signal	$7.62 \pm 3.88$	$2.79 \pm 1.41$
W+t	$1.28 \pm 0.39$	-
$t\bar{t}$	$1.64 \pm 0.35$	$0.45 \pm 0.10$
VZ	$19.90 \pm 4.86$	$7.49 \pm 2.05$
Whf	$0.41 \pm 0.07$	$0.07 \pm 0.01$
Zhf	$150.94 \pm 12.72$	$57.15 \pm 5.81$
Zcl	$2.20 \pm 0.91$	$1.80 \pm 0.76$
Zl	$0.94 \pm 0.67$	$1.01 \pm 0.67$
Data	179	73
Background	$177 \pm 12$	$68 \pm 6$

**Table 7.17:** Post-fit yields in the 2-lepton channel. Combined statistical and systematic uncertainties are shown [190].



**Figure 7.6:** The  $m_J$  post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left)  $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$  and (right)  $p_T^V \geq 400 \text{ GeV}$ . The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions are shown as filled histograms. The Higgs boson signal ( $m_H = 125 \text{ GeV}$ ) is shown as a filled histogram and is normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and as an unstacked unfilled histogram, scaled by the SM prediction times a factor of two. The size of the combined on the sum of the fitted signal and background is shown in the hatched band. The highest bin contains the overflow [142].



**Figure 7.7:** The  $m_J$  post-fit distributions in the  $t\bar{t}$  control region for (top) the 0-lepton channel and the 1-lepton channel for  $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$  and (bottom) the 0-lepton channel and the 1-lepton channel for  $p_T^V > 400 \text{ GeV}$ . The background contributions after the likelihood fit are shown as filled histograms. The Higgs boson signal ( $m_h = 125 \text{ GeV}$ ) is shown as a filled histogram on top of the fitted backgrounds normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and unstacked as an unfilled histogram, scaled by the SM prediction times a factor of 2. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The highest bin in the distributions contains the overflow [142].

### 7.5.2 Observed Signal Strength & Significance

The measured signal strength is computed as the ratio between the measured signal yield to the prediction from the SM. The combined result for all three lepton channels and all analysis regions is given for  $\mu_{VH}^{bb}$  in Eq. (7.8), and for  $\mu_{VZ}^{bb}$  in Eq. (7.9). Both results include a full breakdown of the systematic and statistical uncertainties.

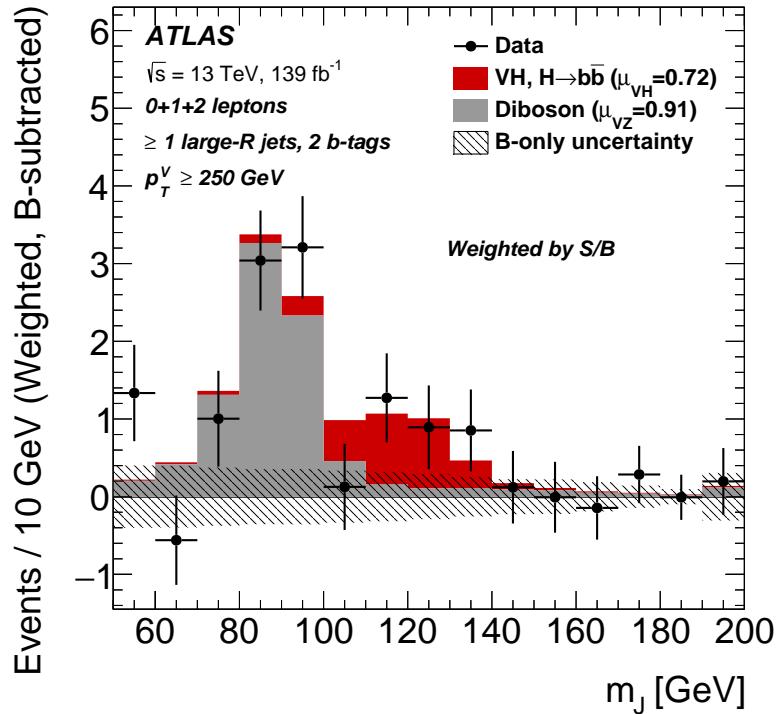
$$\mu_{VH}^{bb} = 0.72_{-0.36}^{+0.39} = 0.72_{-0.28}^{+0.29}(\text{stat.})_{-0.22}^{+0.26}(\text{syst.}) \quad (7.8)$$

$$\mu_{VZ}^{bb} = 0.91_{-0.23}^{+0.29} = 0.91 \pm 0.15(\text{stat.})_{-0.17}^{+0.24}(\text{syst.}) \quad (7.9)$$

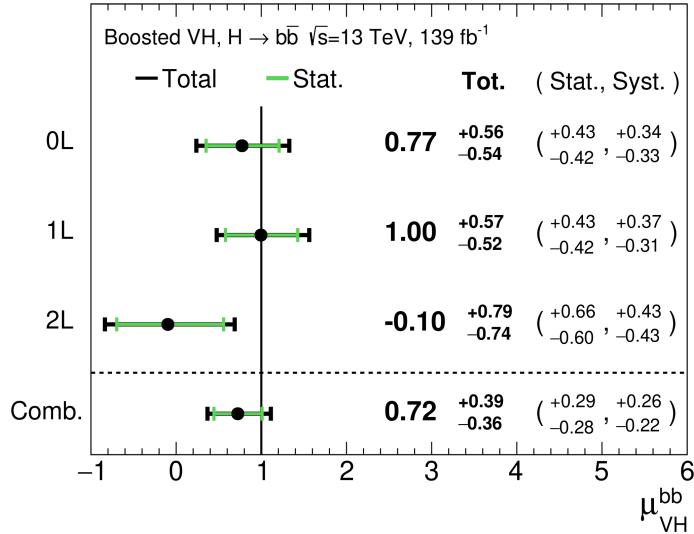
The results for  $\mu_{VH}^{bb}$  and  $\mu_{VZ}^{bb}$  are consistent with the expectation from the SM. The  $\mu_{VH}^{bb}$  measurement is dominated by statistical uncertainty, while the  $\mu_{VZ}^{bb}$  measurement is dominated by systematic sources of uncertainty. The measured signal strength for  $\mu_{VZ}^{bb}$  corresponds to an observed significance of 2.1 standard deviations, with an expected (median) significance given the SM prediction of 2.7 standard deviations. The diboson observed (expected) signal strength significance is 5.4 (5.7). These results are summarised in Fig. 7.8, which shows the weighted and background-subtracted  $m_J$  distribution. A clear signal excess is visible around the Higgs mass of  $m_H = 125$  GeV.

### Compatibility Studies

Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting  $\mu_{VH}^{bb}$  into three separate POIs, one for each channel. A simultaneous fit to the channel specific signal strengths can then be performed, which allows a comparison of the contributions from each channel. Fig. 7.9 compares the best-fit signal strengths. The 0- and 1-lepton channels show a signal strength which is consistent with the SM prediction, while the 2-lepton channel shows a small deviation within the  $1\sigma$  uncertainty. Overall, good compatibility is observed via a  $\chi^2$  test with a corresponding  $p$ -value of 49%.



**Figure 7.8:**  $m_J$  distribution in data after subtraction of all backgrounds except for the  $WZ$  and  $ZZ$  diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected contribution of the associated  $WH$  and  $ZH$  production of a SM Higgs boson with  $m_H = 125 \text{ GeV}$  is shown scaled by the measured combined signal strength ( $\mu_{VH}^{bb} = 0.72$ ). The diboson contribution is normalised to its best-fit value of  $\mu_{VZ}^{bb} = 0.91$ . The size of the combined statistical and systematic uncertainty is indicated by the hatched band. This error band is computed from a full signal-plus-background fit including all the systematic uncertainties defined in Section 7.3, except for the  $VH/VZ$  experimental and theory uncertainties [142].



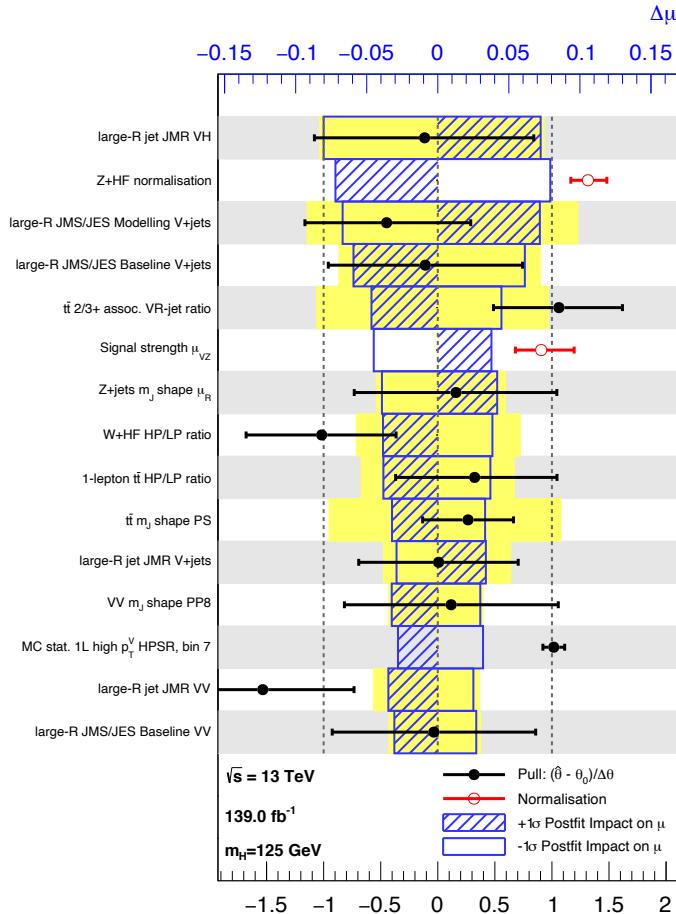
**Figure 7.9:** Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels fit separately) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a  $\chi^2$  difference test and results in a p-value of 49% [142].

### 7.5.3 Impact of Systematics

The impact of systematic uncertainties on the final fitted value  $\hat{\mu}^{bb}$  can be studied using the NP rankings, and the uncertainty breakdown.

Fig. 7.10 shows the NP ranking, which is used to visualise which NPs have the largest impact on the sensitivity to the fitted POI. To obtain the ranking, a likelihood scan is performed for each NP  $\theta_j$ . First, an unconditional fit is used to determine  $\hat{\theta}_j$ . The NP is then fixed to its post-fit value varied by  $\pm 1\sigma$ , the fit is repeated and the difference between the best-fit value of the POI,  $\Delta\hat{\mu}_{VH}^{bb}$ , is calculated, and used to rank the NPs. In addition, the pulls and constraints for the highest ranked NPs are also shown.

The experimental uncertainty on the signal large- $R$  jet mass resolution (JMR) has the largest impact. JMR and jet energy scale (JES) uncertainties also have large impacts for the  $V+jets$  and for the diboson backgrounds. The freely-floating  $Z+hf$  normalisation is the second highest ranked NP, and is heavily constrained by the fit. The  $VZ$  POI  $\mu_{VZ}^{bb}$  is also a significant NP when considering the primary  $\mu_{VH}^{bb}$  measurement.



**Figure 7.10:** Impact of systematic uncertainties on the fitted  $VH$  signal-strength parameter  $\hat{\mu}_{VH}^{bb}$  sorted in decreasing order. The boxes show the variations of  $\hat{\mu}$ , referring to the top  $x$ -axis, when fixing the corresponding individual nuisance parameter to its post-fit value modified upwards or downwards by its post-fit uncertainty, i.e.  $\hat{\theta} \pm \sigma_{\hat{\theta}}$ , and repeating the fit. The impact of up- and down-variations can be distinguished via the dashed and plane box fillings. The yellow boxes show the pre-fit impact (top  $x$ -axis) by varying each nuisance parameter by  $\pm 1$ . The filled circles show the deviation of the fitted value for each nuisance parameter,  $\hat{\theta}$ , from their nominal input value  $\theta_0$  expressed in standard deviations with respect to their nominal uncertainties  $\Delta\theta$  (bottom  $x$ -axis). The error bars show the post-fit uncertainties on  $\hat{\theta}$  with respect to their nominal uncertainties. The open circles show the fitted values and uncertainties of the normalization parameters that are freely floating in the fit. Pre-fit, these parameters have a value of one [190].

The NP ranking highlights individual NPs which have a large impact on the POI measurement sensitivity. Complementary information is provided at a higher level by considering the overall impact of different groups of systematics. The groups are constructed from NPs which have similar physical origins. The impact of each group is calculated by running a fit with all the NPs in the given group fixed to their nominal values. The uncertainty on the POI extracted from this fit is subtracted in quadrature from the uncertainty on the POI from the nominal fit, and the resulting values are provided as the impact for each group. The total systematic impact is the difference in quadrature between the nominal uncertainty on  $\mu_{VH}^{bb}$  and the estimated impact for the combined statistical uncertainties. The “data stat only” group fixes all NPs at their nominal value, while the total statistical impact fixes all NPs except floating normalisations. The floating normalisations group fixes only the NPs associated with normalisation which are left floating in the fit.

The full breakdown for the observed impact of uncertainties on the  $\mu_{VH}^{bb}$  signal strength is provided in Table 7.18. The uncertainty on  $\mu_{VH}^{bb}$  is dominated by combined statistical effects (0.28), although the combined impact of systematics (0.24) is of a comparable size. The signal largest group is the data stat uncertainty (0.25), demonstrating that the analysis would benefit from an increased integrated luminosity or improved efficiency to select signal events (recall from Section 7.1.2 the signal efficiency is in the range of 10%). Of the experimental systematic sources of uncertainty, the dominant impact is from the experimental uncertainties associated with the simulation of large- $R$  jets (0.13). Other experimental sources of uncertainty are small in comparison. Modelling uncertainties also have a large contribution to the overall systematic uncertainty. The biggest contribution to the overall uncertainty is the combined statistical uncertainty on the simulated samples (0.09). Out of the backgrounds, the  $W+jets$  and  $Z+jets$  have the highest (0.06) and second-highest (0.05) impact respectively.

#### 7.5.4 STXS Interpretation

The Simplified Template Cross Sections (STXS) framework provides a common categorisation of candidate Higgs boson events according to certain truth-level properties of the production mode under study [28, 203]. The STXS framework is

Source of uncertainty	Signed impact	Avg. impact
Total	+0.388 / -0.356	0.372
Statistical	+0.286 / -0.280	0.283
↔ Data stat only	+0.251 / -0.245	0.248
↔ Floating normalisations	+0.096 / -0.092	0.094
Systematic	+0.261 / -0.219	0.240
<hr/>		
Experimental uncertainties		
Small- $R$ jets	+0.041 / -0.034	0.038
Large- $R$ jets	+0.161 / -0.105	0.133
$E_T^{\text{miss}}$	+0.008 / -0.007	0.007
Leptons	+0.013 / -0.007	0.010
$b$ -tagging	$b$ -jets	+0.028 / -0.004
	$c$ -jets	+0.012 / -0.011
	light-flavour jets	+0.009 / -0.007
	extrapolation	+0.004 / -0.005
Pile-up	+0.001 / -0.002	0.001
Luminosity	+0.019 / -0.007	0.013
<hr/>		
Theoretical and modelling uncertainties		
Signal	+0.073 / -0.026	0.050
Backgrounds	+0.106 / -0.095	0.100
↔ $Z$ + jets	+0.049 / -0.047	0.048
↔ $W$ + jets	+0.059 / -0.056	0.058
↔ $t\bar{t}$	+0.037 / -0.032	0.035
↔ Single top quark	+0.031 / -0.023	0.027
↔ Diboson	+0.034 / -0.029	0.032
↔ Multijet	+0.009 / -0.009	0.009
↔ MC statistical	+0.091 / -0.092	0.092

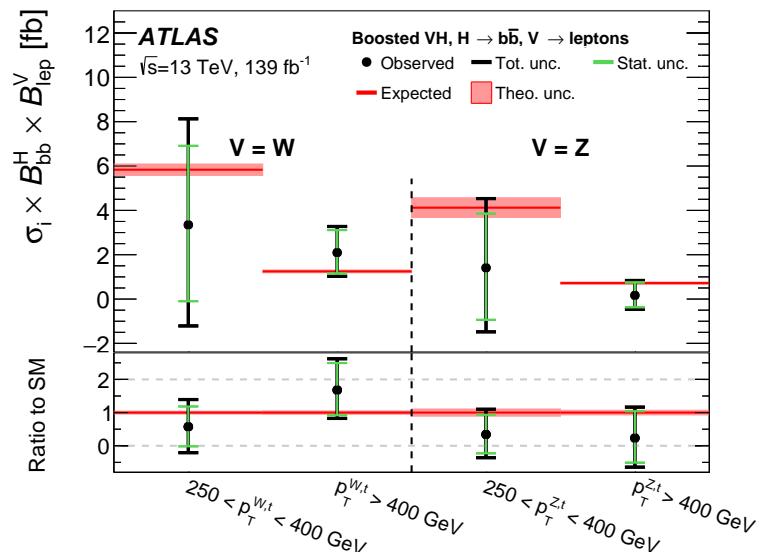
**Table 7.18:** Breakdown of the absolute contributions to the uncertainty on the signal strength  $\mu_{VH}^{bb}$  obtained from the (1+1)-POI fit. The average impact represents the average between the positive and negative uncertainties on  $\mu_{VH}^{bb}$ . The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations [190].

designed to be independent of the decay mode of the Higgs boson, and is therefore well suited to the combination of measurements between different decay channels and experiments.

The STXS cross sections are independently measured for the  $ZH$  and  $WH$  production modes following the approach described in [193]. For each production mode, two bins in the truth vector boson transverse momentum  $p_T^{V,t}$  are considered,  $250 \text{ GeV} < p_T^{V,t} < 400 \text{ GeV}$  and  $p_T^{V,t} \geq 400 \text{ GeV}$ , leading to four independent analysis regions. Events from the simulated signal samples are categorised into the regions and used to estimate the expected cross section times branching ratio  $\sigma \times B$  in each region, where

$$B = B(H \rightarrow b\bar{b}) \times B(V \rightarrow \text{leptons}), \quad (7.10)$$

A simultaneous fit of the four cross section times branching ratios is performed. The uncertainties described in Section 7.3 are reused for the STXS fit, with the exception of the theoretical uncertainties on the signal cross section and branching ratios. The result from the fit is shown in Section 7.5.4 and compared with the expected prediction from the SM. The expected and observed results agree within the given uncertainties.



**Figure 7.11:** Measured  $VH$  simplified template cross sections times the  $H \rightarrow b\bar{b}$  and  $V \rightarrow \text{leptons}$  branching fractions in the medium and high  $p_T^{V,t}$  bins [142].

## 7.6 Conclusion

The analysis of the associated production of vector bosons with boosted Higgs bosons decaying to a pair of  $b$ -quarks using large- $R$  jets is presented. The Higgs candidate is reconstructed as a large- $R$  jet in order to improve sensitivity in the boosted regime in which the Higgs decay products are significantly collimated. The analysis is performed using  $139\text{ fb}^{-1}$  of proton–proton collision data at  $\sqrt{s} = 13\text{ TeV}$  collected throughout the duration of Run 2 of the LHC.

In comparison with the null hypothesis, the Standard Model (SM)  $VH$ ,  $H \rightarrow b\bar{b}$  process is found to have an observed significance of 2.1 standard deviations, whereas the corresponding expected significance is 2.7 standard deviations. This is in contrast to the resolved analysis, which rejected the background-only hypothesis with an observed (expected) significance of 6.7 (6.7) [141]. The  $VH$ ,  $H \rightarrow b\bar{b}$  process is measured simultaneously with the diboson  $VZ$ ,  $Z \rightarrow b\bar{b}$  process, which provide a cross-check for the main analysis. The observed (expected) significance for the diboson process is 5.4 (5.7).

The results are interpreted in the context of the STXS framework. The cross sections for the  $WH$  and  $ZH$  processes are measured in two  $p_T^V$  bins, and are found to agree with the SM prediction within the given uncertainties. At the time of writing, the results are the most precise measurements of the  $WH$  and  $ZH$ ,  $H \rightarrow b\bar{b}$  cross sections in the high  $p_T^V$  regime.

The statistical and systematic sources of uncertainty contribute a similar amount to the overall uncertainty on the result. This analysis would therefore likely benefit greatly from the improved  $b$ -tagging efficiency at high- $p_T$  enabled by GN1 as discussed in Chapter 6, due to the associated reduction in statistical uncertainty provided by the increased number of events used in the analysis.

The large- $R$  jet mass resolution is found to be the dominant source of systematic uncertainty on the  $\mu_{VH}^{bb}$  measurement. An improved method of reconstructing the large- $R$  jet mass, for example by using a machine learning based regression approach, possibly as an additional auxiliary task to GN1, could reduce the systematic uncertainty on the  $\mu_{VH}^{bb}$  measurement. Statistical uncertainty could be reduced by increasing the integrated luminosity used to perform the analysis by combining with Run 3 data.

# Chapter 8

## Conclusion

### 8.1 Summary

The current understanding of particle physics contains many unanswered questions, and improving our understanding of the Standard Model is a promising way to attempt to answer some of them. One of the key particles which may enhance this understanding is the Higgs boson, which was first observed only a decade ago and remains under intense scrutiny at the LHC. Given it's propensity to decay to heavy flavour  $b$ -quarks, reconstructing and identifying  $b$ -jets is of crucial importance to improving our understanding in this area. As discussed in Chapter 4, this task becomes increasingly difficult at high transverse momenta.

One of the effects that hampered tracking and  $b$ -tagging performance at high- $p_{\mathrm{T}}$  was identified to be the increased rate of fake tracks. To address this issue, an algorithm was developed which is able to successfully identify fake tracks within jets 45% of the time, with a minimal loss of signal tracks of 1.2%. Removal of such tracks was found to improve the light-jet mistagging rate of the SV1 and JetFitter algorithms by up to 20% at high transverse momentum.

A novel approach to  $b$ -tagging, GN1 was also developed using a Graph Neural Network (GNN) architecture. The model is encouraged to learn the topology of the jet through vertexing and track classification auxiliary tasks. As a single end-to-end trained model, GN1 simplifies the complexity of the flavour tagging pipeline and is able to achieve superior performance to the current state-of-the-art algorithms, which rely on a two-tiered approach. Compared with DL1r, GN1 improves the light-jet rejection

by a factor of  $\sim 1.8$  for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250 \text{ GeV}$  at the 70%  $b$ -jet WP and by a factor of  $\sim 6$  for jets in the  $Z'$  sample with  $250 < p_T < 5000 \text{ GeV}$  for a corresponding  $b$ -jet efficiency of 30%. GN1 also demonstrates a significant improvement in the discrimination between  $b$ - and  $c$ -jets, which also contributes to improved  $c$ -tagging performance. In the high- $p_T$  regime, GN1 improves the  $b$ -jet tagging efficiency by 100% for a fixed light-jet rejection of 100. Initial validation of the model in data has been performed. The level of agreement between the simulation and data that is observed is similar to previous flavour tagging algorithms. GN1 has been successfully deployed in the ATLAS High Level Trigger, and shows promising performance when trained on high pile-up samples corresponding to the HL-LHC conditions. Ultimately the improved jet tagging performance enabled by the new algorithm will have a large impact across a broad spectrum of the ATLAS physics programme.

This thesis demonstrates that even with suboptimal track reconstruction in this regime, it is possible to make algorithmic advancements to the flavour tagging pipeline to improve the identification of  $b$ -jets. This work has impacts for any analysis which relies on the identification of  $b$ -jets, including those which are sensitive to the Higgs boson.

Analysis of  $VH$ ,  $H \rightarrow b\bar{b}$  events was also carried out with  $139 \text{ fb}^{-1}$  of Run 2 ATLAS at  $\sqrt{s} = 13 \text{ TeV}$ . Various background modelling uncertainties were derived and investigations into the fit model were carried out. The analysis observed a signal strength of  $\mu_{VH}^{bb} = 0.72^{+0.39}_{-0.36} = 0.72^{+0.29}_{-0.28}(\text{stat.})^{+0.26}_{-0.22}(\text{syst.})$  corresponding to an observed (expected) significance of  $2.1\sigma$  ( $2.7\sigma$ ). The result was validated using a simultaneous fit to the  $VZ$ ,  $Z \rightarrow b\bar{b}$  process, which acts as a cross check to validate the primary analysis. The results of the analysis are the most precise measurements available in the high- $p_T$  for the  $VH$ ,  $H \rightarrow b\bar{b}$  process. The high- $p_T$  region is of particular interest as it is a region of phase space with good sensitivity to new physics.

## 8.2 Future Work

Additional algorithmic improvements are likely to yield further improved flavour tagging performance. Aside from these, large improvements to the flavour tagging performance at high- $p_T$  will be possible if the  $b$ -hadron decay track reconstruction

efficiency and accuracy is improved. Investigations into such improvements, for example loosening the track reconstruction requirements in high- $p_{\mathrm{T}}$  environments, are currently ongoing.

At the moment only the tracks from the Inner Detector and kinematic information about the jet are provided as inputs to the tagging algorithms. In Chapter 6 it was shown that the addition of a simple track-level variable corresponding to the ID of the reconstruction lepton to the model improved tagging performance. However there is still untapped potential in the form of additional information from the full parameters of the reconstructed leptons (making full use of the Calorimeters and Muon Spectrometer), the calorimeter clusters, and even the individual clusters which are used to reconstruct tracks. Providing such additional inputs to the model is likely to complement the information provided by the tracks and further improve performance.

Additional auxiliary training objectives may yield improved performance and also help to add to the explainability of the model. Regression of jet-level quantities such as the transverse momentum and mass, in addition to the truth  $b$ -hadron decay length are promising targets.

The GN1 architecture can be easily optimised for new use cases and topologies, as demonstrated by the studies described in Section 6.7. Other opportunities include a model with only cluster-based inputs, which could be used for a fast trigger preselection on jets without the need to run the computationally expensive tracking algorithms, or improved primary vertexing and pile-up jet tagging algorithms.

For an improved analysis of the  $VH$ ,  $H \rightarrow b\bar{b}$  process, the following considerations could be taken into account. Firstly, the addition of Run 3 data will provide a significant increase in statistics and a corresponding reduction in statistical uncertainties. Improved  $b$ -tagging, enabled by the GN1 model, will also improve the sensitivity of the analysis through improvements in the signal-to-background ratio. A dedicated  $X \rightarrow bb$  version of GN1 which is trained to identify the flavour of large- $R$  jets directly would also be of benefit. Improvements in signal-to-background ratio could also be further improved through the use of a dedicated algorithm to select signal events, rather than relying on a series of selection cuts. Finally, the dominant systematic uncertainties relating to the modelling of large- $R$  jets could be reduced via improved reconstruction and calibration techniques.

# Bibliography

- [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- [2] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al., *Secondary vertex finding in jets with neural networks*, *The European Physical Journal C* 81 (2021) .
- [3] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-027 (2022).
- [4] M. K. Gaillard, P. D. Grannis and F. J. Sciulli, *The Standard model of particle physics*, *Rev. Mod. Phys.* 71 (1999) pp. S96–S111 [[hep-ph/9812285](#)].
- [5] S. F. Novaes, *Standard model: An Introduction*, [[hep-ph/0001283](#)].
- [6] L. Morel, Z. Yao, P. Cladé and S. Guellati-Khélifa, *Determination of the fine-structure constant with an accuracy of 81 parts per trillion*, *Nature* 588 (2020) pp. 61–65.
- [7] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al., *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* 606 (2022) pp. 479–483.
- [8] CDF collaboration, *Observation of top quark production in  $\bar{p}p$  collisions*, *Phys. Rev. Lett.* 74 (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- [9] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* 74 (1995) pp. 2632–2637 [[hep-ex/9503003](#)].
- [10] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* 39 (1977) pp. 252–255.
- [11] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $\sqrt{s} = 540$  GeV*, *Phys. Lett. B* 122

- (1983) pp. 103–116.
- [12] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) pp. 218–224 [[hep-ex/0012035](#)].
  - [13] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
  - [14] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) pp. 508–509.
  - [15] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
  - [16] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) p. 1 [[1207.7214](#)].
  - [17] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
  - [18] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) p. 083C01.
  - [19] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
  - [20] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) pp. 579–588.
  - [21] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
  - [22] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, Conf. Proc. C680519 (1968) pp. 367–377.
  - [23] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*, *Phys. Rev.* **104** (1956) pp. 254–258.
  - [24] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hopper and R. P. Hudson, *Experimental test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
  - [25] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of conservation of parity and charge conjugation in meson decays: the magnetic moment*

- of the free muon, *Phys. Rev.* 105 (1957) pp. 1415–1417.*
- [26] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity, *Phys. Rev.* 117 (1960) pp. 648–663.*
- [27] J. Goldstone, *Field theories with «superconductor» solutions, *Il Nuovo Cimento (1955-1965)* 19 (1961) pp. 154–164.*
- [28] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [[1610.07922](#)].
- [29] ATLAS Collaboration, *Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector*, [ATLAS-CONF-2018-036](#) (2018).
- [30] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks, *Phys. Rev. Lett.* 121 (2018) p. 121801* [[1808.08242](#)].
- [31] L. Evans and P. Bryant, *LHC Machine*, *JINST* 3 (2008) p. S08001.
- [32] The ALICE Collaboration, *The ALICE experiment at the CERN LHC, *Journal of Instrumentation* 3 (2008) pp. S08002–S08002.*
- [33] CMS Collaboration, *The CMS experiment at the CERN LHC, *JINST* 3 (2008) p. S08004.*
- [34] The LHCb Collaboration, *The LHCb detector at the LHC, *Journal of Instrumentation* 3 (2008) pp. S08005–S08005.*
- [35] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider, *JINST* 3 (2008) p. S08003.*
- [36] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- [37] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [38] CERN, *The accelerator complex*, (2012). <http://public.web.cern.ch/public/en/research/AccelComplex-en.html>.
- [39] G. C. Strong, *On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy*

- physics use case*, [[2002.01427](#)].
- [40] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022).  
<https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoverview/>.
- [41] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, *JINST* 11 (2016) p. P04008 [[1512.01094](#)].
- [42] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 13 \text{ TeV}$  using the ATLAS detector at the LHC*, *ATLAS-CONF-2019-021* (2019).
- [43] G. Avoni, M. Bruschi, G. Cabras, D. Caforio, N. Dehghanian, A. Floderus et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, *JINST* 13 (2018) p. P07017.
- [44] ATLAS Collaboration, *ATLAS data quality operations and performance for 2015–2018 data-taking*, *JINST* 15 (2020) p. P04003 [[1911.04632](#)].
- [45] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design Report, Volume 1*, ATLAS-TDR-14; CERN-LHCC-99-014, Geneva (1999).  
<https://cds.cern.ch/record/391176>.
- [46] ATLAS Collaboration, *Status of the ATLAS Experiment*, Tech. Rep., Geneva (2010).  
<https://cds.cern.ch/record/1237407>.
- [47] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).  
<https://cds.cern.ch/record/2285585>.
- [48] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).  
<https://cds.cern.ch/record/2257755>.
- [49] ATLAS Collaboration, *The inner detector*, (2022).  
<https://atlas.cern/Discover/Detector/Inner-Detector>.
- [50] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).  
<https://cds.cern.ch/record/1291633>.
- [51] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* 13 (2018) p. T05008 [[1803.00844](#)].
- [52] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor*

- tracker*, *JINST* 9 (2014) p. P08009 [[1404.7473](#)].
- [53] ATLAS Collaboration, *Calorimeter*, (2022).  
<https://atlas.cern/Discover/Detector/Calorimeter>.
- [54] F. Cavallari, *Performance of calorimeters at the LHC*, *J. Phys. Conf. Ser.* 293 (2011) p. 012001.
- [55] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, *JINST* 14 (2019) p. P03017 [[1812.03848](#)].
- [56] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13\text{ TeV}$* , *Eur. Phys. J. C* 76 (2016) p. 292 [[1603.05598](#)].
- [57] ATLAS Collaboration, *Muon spectrometer*, (2022).  
<https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- [58] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* 77 (2017) p. 317 [[1611.09661](#)].
- [59] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001 (2021).
- [60] ATLAS Collaboration, *ATLAS Computing Acknowledgements*, Tech. Rep.  
<https://cds.cern.ch/record/2717821>.
- [61] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- [62] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense Environments*, ATL-PHYS-PUB-2015-006 (2015).
- [63] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, *Eur. Phys. J. C* 77 (2017) p. 673 [[1704.07983](#)].
- [64] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, *JINST* 9 (2014) p. P09009 [[1406.7690](#)].
- [65] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS Experiment*, Ph.D. Thesis, Innsbruck U., 2013.  
<https://cds.cern.ch/record/1625231>.

- [66] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332 [[1611.10235](#)].
- [67] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at  $\sqrt{s} = 13$  TeV*, **ATL-PHYS-PUB-2015-026** (2015).
- [68] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with  $t\bar{t}$  events in  $pp$  collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79** (2019) p. 970 [[1907.05120](#)].
- [69] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, **ATL-PHYS-PUB-2017-011** (2017).
- [70] M. Cacciari, G. P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) p. 063 [[0802.1189](#)].
- [71] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur.Phys.J. C* **72** (2012) p. 1896 [[1111.6097](#)].
- [72] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) p. 005 [[0802.1188](#)].
- [73] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, **ATLAS-CONF-2014-018** (2014).
- [74] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- [75] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].
- [76] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- [77] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008) p. 242001 [[0802.2470](#)].
- [78] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) p. 084 [[0912.1342](#)].
- [79] ATLAS Collaboration, *Performance of jet substructure techniques for large- $R$  jets in*

- proton–proton collisions at  $\sqrt{s} = 7 \text{ TeV}$  using the ATLAS detector,* *JHEP* 09 (2013) p. 076 [[1306.4945](#)].
- [80] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* 48 (1993) pp. 3160–3166 [[hep-ph/9305266](#)].
- [81] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* 79 (2019) p. 135 [[1807.09477](#)].
- [82] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* 06 (2009) p. 059 [[0903.0392](#)].
- [83] ATLAS Collaboration, *Variable Radius, Exclusive- $k_T$ , and Center-of-Mass Subjet Reconstruction for Higgs( $\rightarrow b\bar{b}$ ) Tagging in ATLAS*, [ATL-PHYS-PUB-2017-010](#) (2017).
- [84] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, [ATLAS-CONF-2012-047](#) (2012).
- [85] ATLAS Collaboration, *Electron and photon reconstruction and performance in ATLAS using a dynamical, topological cell clustering-based approach*, [ATL-PHYS-PUB-2017-022](#) (2017).
- [86] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, *JINST* 14 (2019) p. P12006 [[1908.00005](#)].
- [87] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, [ATLAS-CONF-2016-024](#) (2016).
- [88] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 13 \text{ TeV}$* , *Eur. Phys. J. C* 78 (2018) p. 903 [[1802.08168](#)].
- [89] ATLAS Collaboration, *Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector*, *Phys. Lett. B* 786 (2018) p. 59 [[1808.08238](#)].
- [90] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* 784 (2018) p. 173 [[1806.00425](#)].
- [91] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*

- using  $139\text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector, *JHEP* 03 (2020) p. 145 [[1910.08447](#)].
- [92] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* 15S1 (2000) pp. 577–606 [[hep-ph/9912292](#)].
- [93] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* 98 (2018) p. 030001.
- [94] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high  $p_T$  jets*, ATL-PHYS-PUB-2014-008 (2014).
- [95] Nazar Bartosik, *Diagram showing the common principle of identification of jets initiated by b-hadron decays*, (2022).  
[https://en.m.wikipedia.org/wiki/File:B-tagging\\_diagram.png](https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png).
- [96] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003 (2017).
- [97] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014 (2020).
- [98] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022 (2015).
- [99] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017).
- [100] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, [[2211.16345](#)].
- [101] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*, *Journal of High Energy Physics* 2004 (2004) p. 040–040.
- [102] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* 2007 (2007) p. 126–126.
- [103] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton shower simulations: the powheg method*, *Journal of High Energy Physics* 2007 (2007) p. 070–070.
- [104] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*

- calculations in shower monte carlo programs: the powheg box,* *Journal of High Energy Physics* 2010 (2010) .
- [105] NNPDF collaboration, *Parton distributions for the LHC run II,* *JHEP* 04 (2015) p. 040 [[1410.8849](#)].
- [106] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016,* *ATL-PHYS-PUB-2016-020* (2016).
- [107] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2,* *Comput. Phys. Commun.* 191 (2015) p. 159 [[1410.3012](#)].
- [108] ATLAS Collaboration, *Study of top-quark pair modelling and uncertainties using ATLAS measurements at  $\sqrt{s} = 13$  TeV,* *ATL-PHYS-PUB-2020-023* (2020).
- [109] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data,* *ATL-PHYS-PUB-2014-021* (2014).
- [110] R. D. Ball et al., *Parton distributions with LHC data,* *Nucl. Phys. B* 867 (2013) p. 244 [[1207.1303](#)].
- [111] D. J. Lange, *The EvtGen particle decay simulation package,* *Nucl. Instrum. Meth. A* 462 (2001) p. 152.
- [112] ATLAS Collaboration, *The ATLAS Simulation Infrastructure,* *Eur. Phys. J. C* 70 (2010) p. 823 [[1005.4568](#)].
- [113] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit,* *Nucl. Instrum. Meth. A* 506 (2003) p. 250.
- [114] ATLAS Collaboration, *Tracking efficiency studies in dense environments,* (2022). <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2022-03/>.
- [115] S. Adorni Braccesi Chiassi, *Vector Boson Tagging in the Context of the Search for New Physics in the Fully Hadronic Final State,* Ph.D. Thesis, Geneva U., 2021. [10.13097/archive-ouverte/unige:158558](https://doi.org/10.13097/archive-ouverte/unige:158558).
- [116] E. Boos et al., *Generic user process interface for event generators,* [[hep-ph/0109068](#)].
- [117] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A standard format for Les Houches Event Files,* *Computer Physics Communications* 176 (2007) pp. 300–304.

- [118] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are universal approximators*, *Neural Networks* 2 (1989) pp. 359–366.
- [119] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by back-propagating errors*, *nature* 323 (1986) pp. 533–536.
- [120] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, *The bulletin of mathematical biophysics* 5 (1943) pp. 115–133.
- [121] J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- [122] A. F. Agarap, *Deep Learning using Rectified Linear Units (ReLU)*, *arXiv e-prints* (2018) p. arXiv:1803.08375 [1803.08375].
- [123] S. Elfwing, E. Uchibe and K. Doya, *Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning*, *arXiv e-prints* (2017) p. arXiv:1702.03118 [1702.03118].
- [124] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- [125] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv e-prints* (2014) [1412.6980].
- [126] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al., *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. arXiv:2002.08772 [2002.08772].
- [127] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*, *arXiv e-prints* (2018) p. arXiv:1806.01261 [1806.01261].
- [128] J. Shlomi, P. Battaglia and J.-R. Vlimant, *Graph neural networks in particle physics*, *Machine Learning: Science and Technology* 2 (2021) p. 021001.
- [129] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola, *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [1703.06114].
- [130] ATLAS Collaboration, *Muon reconstruction performance in early  $\sqrt{s} = 13$  TeV data*, *ATL-PHYS-PUB-2015-037* (2015).
- [131] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*

- experiment using the 2015 and 2016 LHC proton–proton collision data at  $\sqrt{s} = 13\text{ TeV}$ , *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].*
- [132] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, arXiv e-prints (2021) [[2007.08294](#)].
- [133] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*, arXiv e-prints (2021) [[2105.14491](#)].
- [134] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al., *Attention Is All You Need*, arXiv e-prints (2017) [[1706.03762](#)].
- [135] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv e-prints (2014) [[1412.6980](#)].
- [136] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).  
<https://github.com/onnx/onnx>.
- [137] ATLAS Collaboration, *Performance of Run 3 HLT b-tagging with fast tracking, Public b-Jet Trigger Plots for Collision Data*, (2022).  
[https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults#Performance\\_of\\_Run\\_3\\_HLT\\_b\\_taggi](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults#Performance_of_Run_3_HLT_b_taggi).
- [138] ATLAS Collaboration, *Neural Network Jet Flavour Tagging with the Upgraded ATLAS Inner Tracker Detector at the High-Luminosity LHC*, ATL-PHYS-PUB-2022-047 (2022).
- [139] ATLAS Collaboration, *Jet flavour tagging with GN1 and DL1d. generator dependence, run 2 and run 3 data agreement studies*, (2023).  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/>.
- [140] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].
- [141] ATLAS Collaboration, *Measurements of WH and ZH production in the  $H \rightarrow b\bar{b}$  decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) p. 178 [[2007.02873](#)].
- [142] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector*, *Phys. Lett. B* **816** (2021) p. 136204 [[2008.02508](#)].

- [143] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039 [[1512.02572](#)].
- [144] ATLAS Collaboration, *ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta*, **ATL-PHYS-PUB-2009-088** (2009).
- [145] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using  $t\bar{t}$  events at  $\sqrt{s} = 13$  TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- [146] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using ATLAS proton–proton collision data at  $\sqrt{s} = 13$  TeV*, **ATLAS-CONF-2018-006** (2018).
- [147] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in  $t\bar{t}$  events using a likelihood approach with the ATLAS detector*, **ATLAS-CONF-2018-001** (2018).
- [148] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the ATLAS Detector*, **ATL-PHYS-PUB-2014-013** (2014).
- [149] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) p. 836 [[1906.11005](#)].
- [150] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) p. 043 [[1002.2581](#)].
- [151] ATLAS collaboration, *Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **09** (2014) p. 145 [[1406.3660](#)].
- [152] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al., *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* **72** (2012) p. 1889 [[1111.2034](#)].
- [153] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*, *JHEP* **10** (2012) p. 155 [[1206.3572](#)].
- [154] G. Luisoni, P. Nason, C. Oleari and F. Tramontano,  *$HW^\pm/HZ + 0$  and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, *JHEP* **10** (2013) p. 083 [[1306.2542](#)].

- [155] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, *Phys. Rev. D* **68** (2003) p. 073003 [[hep-ph/0306234](#)].
- [156] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) pp. 149–156 [[hep-ph/0307206](#)].
- [157] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.* **107** (2011) p. 152003 [[1107.1164](#)].
- [158] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) p. 1868 [[1111.0761](#)].
- [159] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated WH production and decay at the LHC*, *JHEP* **04** (2014) p. 039 [[1312.1669](#)].
- [160] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron colliders: The fully differential NNLO QCD calculation*, *Phys. Lett. B* **740** (2015) pp. 51–55 [[1407.4747](#)].
- [161] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson at NNLO*, *JHEP* **06** (2016) p. 179 [[1601.00658](#)].
- [162] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke, *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) p. 078 [[1211.5015](#)].
- [163] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) p. 065 [[1503.01656](#)].
- [164] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) p. 082 [[1410.0217](#)].
- [165] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) p. 023 [[1307.8122](#)].
- [166] O. Brein, R. V. Harlander and T. J. E. Zirke,  *$vh@nnlo$  – Higgs Strahlung at hadron colliders*, *Comput. Phys. Commun.* **184** (2013) pp. 998–1003 [[1210.5347](#)].
- [167] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) p. 126 [[0707.3088](#)].

- [168] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, *Comput. Phys. Commun.* 185 (2014) p. 2930 [[1112.5675](#)].
- [169] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* 09 (2009) p. 111 [[0907.4076](#)].
- [170] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, *Phys. Rev. D* 81 (2010) p. 054028 [[1001.5034](#)].
- [171] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev. D* 83 (2011) p. 091503 [[1103.2792](#)].
- [172] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* 71 (2011) p. 1547 [[1009.2450](#)].
- [173] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a  $W^-$  or  $H^-$* , *Phys. Rev. D* 82 (2010) p. 054018 [[1005.4451](#)].
- [174] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al., *Event generation with SHERPA 1.1*, *JHEP* 02 (2009) p. 007 [[0811.4622](#)].
- [175] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* 7 (2019) p. 034 [[1905.09127](#)].
- [176] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* 108 (2012) p. 111601 [[1111.5206](#)].
- [177] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* 12 (2008) p. 039 [[0808.3674](#)].
- [178] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, *JHEP* 03 (2008) p. 038 [[0709.1027](#)].
- [179] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, *JHEP* 04 (2013) p. 027 [[1207.5030](#)].
- [180] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.* 103 (2009) p. 082001 [[0903.2120](#)].
- [181] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* 43 (2016) p. 023001 [[1510.03865](#)].

- [182] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* 03 (2012) p. 075 [[1112.5142](#)].
- [183] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs strahlung at hadron colliders*, *Comput. Phys. Commun.* 195 (2015) pp. 161–171 [[1412.5390](#)].
- [184] ATLAS Collaboration, *Jet mass and substructure of inclusive jets in  $\sqrt{s} = 7 \text{ TeV}$  pp collisions with the ATLAS experiment*, *JHEP* 05 (2012) p. 128 [[1203.4606](#)].
- [185] ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, [ATLAS-CONF-2016-035](#) (2016).
- [186] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* 117 (2016) p. 182002 [[1606.02625](#)].
- [187] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C*75 (2015) p. 204 [[1412.3989](#)].
- [188] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev. D*93 (2016) p. 033006 [[1506.07443](#)].
- [189] J. K. Anders and M. D’Onofrio, *V+Jets theoretical uncertainties estimation via a parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016). <https://cds.cern.ch/record/2125718>.
- [190] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a pair of b-quarks using large-R jets*, [ATL-COM-PHYS-2019-1125](#) (2019).
- [191] J. Pumplin et al., *New generation of parton distributions with uncertainties from global qcd analysis*, *JHEP* 07 (2002) p. 012 [[0201195](#)].
- [192] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top hadroproduction in association with a W boson*, *JHEP* 07 (2008) p. 029 [[0805.3067](#)].
- [193] ATLAS Collaboration, *Measurement of VH,  $H \rightarrow b\bar{b}$  production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector*, *JHEP* 05 (2019) p. 141 [[1903.04618](#)].

- [194] ATLAS Collaboration, *Evidence for the  $H \rightarrow b\bar{b}$  decay with the ATLAS detector*, *JHEP* 12 (2017) p. 024 [[1708.03299](#)].
- [195] ATLAS Collaboration, *Evaluation of theoretical uncertainties for simplified template cross section measurements of  $V$ -associated production of the Higgs boson*, *ATL-PHYS-PUB-2018-035* (2018).
- [196] J. R. Andersen et al., *Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report*, [[1803.07977](#)].
- [197] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, *CERN-2012-002* (2012) [[1201.3084](#)].
- [198] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, *CERN-2013-004* (2013) [[1307.1347](#)].
- [199] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The roostats project*, [[1009.1003](#)].
- [200] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, [[physics/0306116](#)].
- [201] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, *Comput. Phys. Commun.* 77 (1993) pp. 219–228.
- [202] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* 71 (2011) p. 1554 [[1007.1727](#)].
- [203] J. R. Andersen et al., *Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report*, [[1605.04692](#)].

# List of figures

2.1	The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$ , with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. . . . .	21
2.2	Higgs boson production cross sections as a function of Higgs mass ( $m_H$ ) at $\sqrt{s} = 13$ TeV [28]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, Higgs boson production is dominated by gluon-gluon fusion, vector boson fusion, associated production with vector bosons, and top quark fusion. . . . .	26
2.3	Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV. . . . .	26
2.4	Higgs boson branching ratios as a function of Higgs mass ( $m_H$ ) at $\sqrt{s} = 13$ TeV [28]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, the Higgs predominantly decays to a pair of $b$ -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of $W$ bosons. . . . .	27
3.1	An overview of the CERN accelerator complex [38]. The LHC is fed by a series of accelerators starting with Linac2 (or Linac4 from 2020). Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC. . . . .	30
3.2	The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [39]. . . . .	31

---

3.3 The track parameterisation used at the ATLAS detector. Five coordinates ( $d_0, z_0, \phi, \theta, q/p$ ) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum $\mathbf{p}$ and the transverse momentum $p_T$ (defined in Eq. (3.2)). The basis vectors $e_x$ , $e_y$ and $e_z$ are also shown. Reproduced from Ref. [40]. . . . .	33
3.4 Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [37]. A total of $139\text{ fb}^{-1}$ of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally [44]. . . . .	35
3.5 Average pile-up profiles measured by ATLAS during Run 2 [37]. . . . .	35
3.6 A 3D model of the entire ATLAS detector [46]. Cutouts to the centre of the detector reveal the different sub-detectors which are arranged in concentric layers around the nominal interaction point. . . . .	36
3.7 A 3D model of the ATLAS ID showing the pixel, SCT and TRT sub-detectors [49]. . . . .	37
3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [40]. . . . .	38
3.9 A schematic cross-sectional view of the ATLAS IBL [50]. . . . .	39
3.10 The ATLAS calorimeters [53]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic end-caps. . . . .	41
3.11 The ATLAS muon spectrometer [57]. . . . .	43
3.12 (left) Particles which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [63]. . . . .	47
3.13 A sketch of electron reconstruction using the ATLAS detector [87]. Electron reconstruction makes use of the entire ID and the calorimeters. . . . .	52

---

3.14 The truth $b$ -hadron decay radius $L_{xy}$ as a function of truth transverse momentum $p_T$ for reconstructed $b$ -jets in $Z'$ events (see Section 4.1). Error bars show the standard deviation of the $L_{xy}$ distribution in each $p_T$ bin. The pixel layers are shown in dashed horizontal lines. . . . .	56
3.15 Diagram of a typical $b$ -jet (blue) which has been produced in an event alongside two light jets (grey) [95]. The $b$ -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter $d_0$ is a characteristic property of the trajectories of $b$ -hadron decay products. . . . .	57
3.16 An overview of different low and high level taggers used in ATLAS. The low level taggers are IPxD, SV1 and JetFitter, and RNNIP and DIPS [68, 69, 96, 97]. The outputs of these taggers are fed into the high-level taggers DL1, DL1r and DL1d [99, 100]. . . . .	58
4.1 At lower $p_T$ (left) the decay length of the $b$ -hadron is on average reduced, and the decay tracks are less collimated. At higher $p_T$ (right) the $b$ -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching detector elements. As a result, the ID may be unable to resolve charged depositions from different particles, resulting merged clusters. . . . .	61
4.2 $b$ -hadron decay track reconstruction efficiency as a function of truth $b$ -hadron $p_T$ for jets in the $Z'$ sample [114]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- $p_T$ $b$ -hadrons, the ambiguity solver is overly aggressive in its removal of $b$ -hadron decay tracks. Suggestions for the improvement of the track reconstruction efficiency in this regime by the loosening of cuts in the ambiguity solver are shown in blue and red. . . . .	63
4.3 Average hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the $p_T$ of the reconstructed track for tracks in jets in a $Z'$ sample at $\sqrt{s} = 13$ TeV. Tracks from the weak decay of the $b$ -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. Baseline tracks are those produced in the standard reconstruction described in Section 3.4.1, while pseudotrack represent the ideal performance of the ATLAS detector and are described in Section 3.4.1. . . . .	63

---

4.4 The fraction of IBL (left) and B-layer (right) hits which are shared on $b$ -hadron decay tracks as a function of the production radius of the $b$ -hadron decay product for tracks in jets in a $Z'$ sample at $\sqrt{s} = 13$ TeV. Pseudotrack represent the ideal performance given the ATLAS detector, see Section 3.4.1.	65
4.5 The average number of good (left) and wrong (right) IBL hits as a function of $b$ -hadron $p_T$ for tracks in the $Z'$ sample. The baseline tracking performance (black) and the modified version of the outlier removal procedure (red) are shown. . . . .	67
4.6 (left) $b$ -hadron decay track $d_0$ pulls ( $d_0/s(d_0)$ ) for baseline and modified GX2F tracks in the $Z'$ sample. (right) The absolute value of the $d_0$ pull as a function of the truth $b$ -hadron transverse momentum. . . . .	68
5.1 A diagram displaying the logical flow of a single neuron with three inputs $x_i^j$ . Each input is multiplied by a weight $\theta_j$ , and the resulting values are summed. A bias term $\theta_0$ is added, and the result $z$ is passed to an activation function. Each neuron can be thought of as a logistic regression model. . . . .	72
5.2 The output of several common choices for the activation function $g(z)$ of an artificial neuron. The input $z$ is the output of the dot product between the activation and the weights, plus a bias term. . . . .	73
5.3 Rate of fake tracks as a function of jet transverse momentum (left) and $\Delta R(\text{track}, \text{jet})$ (right) for jets in the $Z'$ sample. The rate of fake tracks increases significantly as a function of $p_T$ , and also increases as the distance to the jet axis decreases. . . . .	76
5.4 The light-jet efficiency of the low level tagger SV1 for jets in the $Z'$ sample with $250 < p_T < 5000$ GeV, as a function of $b$ -jet efficiency. The nominal tracking setup (black) is shown alongside the case where fake tracks which are not from the decay of a $b$ -hadron are removed. The light-jet efficiency is decreased, demonstrating that the presence of fake tracks is detrimental to the algorithm performance. . . . .	77

---

5.5 (left) Normalised histograms of the fake track classification model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a $b$ -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a $b$ -hadron (dashed line). The plots show tracks in the combined $t\bar{t}$ and $Z'$ testing sample. The model is able to successfully discriminate between signal and fake tracks, and shows a very similar performance when looking specifically at tracks from the decay of a $b$ -hadron. . . . .	80
5.6 (left) Normalised histogram of the $b$ -hadron track identification model output separated for tracks from the decay of a $b$ -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line). The plots show tracks in the combined $t\bar{t}$ and $Z'$ testing sample. . . . .	82
5.7 The effect of applying the fake track identification algorithm together with the $b$ -hadron decay track identification on the jet tagging performance of SV1 for jets in the $Z'$ sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet efficiency (black) is compared to two working points of fake track removal, labelled “A” (red) and “B” (orange), which represent two different 2D working points of the track classification tools. Removal of fake tracks based on truth information is shown by the green curves. . . . .	83
6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right). The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, additional truth information from the simulation is now being used as auxiliary training targets for GN1. The solid lines represent reconstructed information, whereas the dashed lines represent truth information [3]. . .	86
6.2 The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [129]. Reproduced from Ref. [127]. . . . .	89

- 
- 6.3 The inputs to GN1 are the two jet features ( $n_{\text{jf}} = 2$ ), and an array of  $n_{\text{tracks}}$ , where each track is described by 21 track features ( $n_{\text{tf}} = 21$ ). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1 [3]. . . . . 94
- 6.4 The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting node representations are used to predict the jet flavour, the track origins, and the track-pair vertex compatibility [3]. . . . . 94
- 6.5 Comparison between the DL1r and GN1  $b$ -tagging discriminant  $D_b$  for jets in the  $t\bar{t}$  sample. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1. The distributions of the different jet flavours have been normalised to unity area [3]. . . . . 99
- 6.6 The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for jets in the  $t\bar{t}$  sample with  $20 < p_{\text{T}} < 250 \text{ GeV}$ . The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $b$ -jet tagging efficiencies usually probed in these regions of phase space [3]. 102
- 6.7 The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for jets in the  $Z'$  sample with  $250 < p_{\text{T}} < 5000 \text{ GeV}$ . The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $b$ -jet tagging efficiencies usually probed in these regions of phase space [3]. 102

---

6.8 The $b$ -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the $Z'$ sample (right) as a function of jet $p_T$ with a fixed light-jet rejection of 100 in each bin. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. GN1 demonstrates improved performance with respect to DL1r across the $p_T$ range shown. Binomial error bands are denoted by the shaded regions [3]. . . . .	103
6.9 The $b$ -jet (left) and light-jet (right) rejections as a function of the $c$ -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The $x$ -axis range is chosen to display the $c$ -jet tagging efficiencies usually probed in these regions of phase space [3]. . . . .	105
6.10 The $b$ -jet (left) and light-jet (right) rejections as a function of the $c$ -jet tagging efficiency for $Z'$ jets with $250 < p_T < 5000$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The $x$ -axis range is chosen to display the $c$ -jet tagging efficiencies usually probed in these regions of phase space [3]. . . . .	105
6.11 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1 [3]. . . . .	107
6.12 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet tagging efficiency for $Z'$ jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1 [3]. . . . .	107

- 6.13 A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a  $b$ -jet. The true and predicted origins of the tracks is shown by the coloured circles along the diagonal. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities  $p_b$ ,  $p_c$  and  $p_u$  are rounded to three significant figures. GN1 improves over SV1 and JetFitter by successfully determining the origins and vertex compatibility of all the tracks in the jet with the exception of the truth OtherSecondary track, which is misidentified as pile-up. . . . . 109
- 6.14 A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a  $b$ -jet. The true and predicted origins of the tracks is shown by the coloured circles along the diagonal. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities  $p_b$ ,  $p_c$  and  $p_u$  are rounded to three significant figures. GN1 improves over SV1 and JetFitter by successfully determining the origins and vertex compatibility of all but one tracks in the jet. . . . . 109
- 6.15 Heavy flavour vertex finding efficiency as a function of jet  $p_T$  for  $b$ -jets in the  $t\bar{t}$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions. . . 112
- 6.16 Heavy flavour vertex finding efficiency as a function of jet  $p_T$  for  $b$ -jets in the  $Z'$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions. . . 112

- 
- 6.17 Vertex finding efficiency for other secondary decays as a function of jet  $p_T$  for light-jets in the  $t\bar{t}$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions. . . . . 113
- 6.18 Vertex finding efficiency as a function of jet  $p_T$  for light-jets in the  $Z'$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions. . . . . 114
- 6.19 ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the  $t\bar{t}$  sample (left) and jets in the  $Z'$  sample (right). The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the Primary and OtherSecondary labels have similarly been combined into a single category. The mean weighted area under the ROC curves (AUC) is similar for both samples [3]. . . . . 116
- 6.20 The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $t\bar{t}$  jets with  $20 < p_T < 250$  GeV, for the looser track selection trainings of GN1. The ratio to the performance of the baseline GN1 model is shown in the bottom panels. A value of  $f_c = 0.05$  is used for all models. Binomial error bands are denoted by the shaded regions. . . . . 118
- 6.21 The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $Z'$  jets with  $250 < p_T < 5000$  GeV, for the looser track selection trainings of GN1. The ratio to the performance of the baseline GN1 model is shown in the bottom panels. A value of  $f_c = 0.05$  is used for all models. Binomial error bands are denoted by the shaded regions. . . . . 118

6.22 The HLT light-jet rejection as a function of the $b$ -jet efficiency jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of mass energy $\sqrt{s} = 13.6$ TeV [137]. The ratio to the performance of the DL1d algorithm [100] is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for b-tagging. . . . .	119
6.23 The $c$ -jet rejection (left) and light-jet rejection (right) for the upgraded HL-LHC ATLAS detector with $\langle \mu \rangle = 200$ as a function of the $b$ -jet efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of mass energy $\sqrt{s} = 14$ TeV [138]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for b-tagging. . . . .	121
6.24 The $c$ -jet rejection (left) and light-jet rejection (right) for the upgraded HL-LHC ATLAS detector with $\langle \mu \rangle = 200$ as a function of the $b$ -jet efficiency for jets in the $Z'$ sample with $250 < p_T < 5000$ GeV for events with a centre of mass energy $\sqrt{s} = 14$ TeV [138]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. . . . .	121
7.1 Diagrams of the signal process (top) and $t\bar{t}$ background (bottom). The object to the right of centre are reconstructed within the large- $R$ jet. For the $t\bar{t}$ background, the large- $R$ jet contains a mistagged $c$ -jet or (less often) a mistagged light-jet. The contribution in the 0-lepton channel results from hadronically decaying $\tau$ lepton, or a electron or muon which is out of the analysis acceptance. For the $t\bar{t}$ background process to be in the signal region, the $b$ -jet on the left hand side must either be not within acceptance, not reconstructed, or not tagged. . . . .	131
7.2 Leading large- $R$ jet mass for the $Z$ and $W+hf$ processes in the HP SR of the 0-lepton channel [190]. The renormalisation scale $\mu_r$ has been varied by a factor of 0.5 (1DOWN) and 2 (1UP). An exponential function is fitted to the ratio between the nominal and alternate samples. . . . .	143
7.3 Leading large- $R$ jet $m_J$ inclusive in $p_T^V$ for the $V+hf$ process modelled using both the SHERPA +PYTHIA8 (blue) and MADGRAPH +PYTHIA8 (red) samples [190]. . . . .	144

---

7.4	Leading large- $R$ jet mass distribution for the combined $WZ$ and $ZZ$ processes, inclusive across all signal regions and lepton channels [190]. The renormalisation scale $\mu_R$ has been varied by a factor of 2 (UP) and 0.5 (DOWN). The red and blue curves show the fitted results of the hyperbolic tangent function. . . . .	146
7.5	Comparison of the shapes of the large- $R$ jet mass $m_J$ between SHERPA (Sh221) (black) and POWHEG+PYTHIA8 (PwPy) (red) samples from the $WZ$ and $ZZ$ processes in high and low purity signal regions, integrated over $p_T^V$ regions and the 0- and 1-lepton channels [190]. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after protection is added in the high mass region. . . . .	147
7.6	The $m_J$ post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$ . The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions are shown as filled histograms. The Higgs boson signal ( $m_H = 125 \text{ GeV}$ ) is shown as a filled histogram and is normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and as an unstacked unfilled histogram, scaled by the SM prediction times a factor of two. The size of the combined on the sum of the fitted signal and background is shown in the hatched band. The highest bin contains the overflow [142]. . . . .	157
7.7	The $m_J$ post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$ . The background contributions after the likelihood fit are shown as filled histograms. The Higgs boson signal ( $m_h = 125 \text{ GeV}$ ) is shown as a filled histogram on top of the fitted backgrounds normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and unstacked as an unfilled histogram, scaled by the SM prediction times a factor of 2. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The highest bin in the distributions contains the overflow [142]. . . . .	158

---

7.8 $m_J$ distribution in data after subtraction of all backgrounds except for the $WZ$ and $ZZ$ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected contribution of the associated $WH$ and $ZH$ production of a SM Higgs boson with $m_H = 125$ GeV is shown scaled by the measured combined signal strength ( $\mu_{VH}^{bb} = 0.72$ ). The diboson contribution is normalised to its best-fit value of $\mu_{VZ}^{bb} = 0.91$ . The size of the combined statistical and systematic uncertainty is indicated by the hatched band. This error band is computed from a full signal-plus-background fit including all the systematic uncertainties defined in Section 7.3, except for the $VH/VZ$ experimental and theory uncertainties [142]. . . . .	160
7.9 Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels fit separately) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a $\chi^2$ difference test and results in a p-value of 49% [142]. . . . .	161
7.10 Impact of systematic uncertainties on the fitted $VH$ signal-strength parameter $\mu_{VH}^{bb}$ sorted in decreasing order. The boxes show the variations of $\hat{\mu}$ , referring to the top $x$ -axis, when fixing the corresponding individual nuisance parameter to its post-fit value modified upwards or downwards by its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$ , and repeating the fit. The impact of up- and down-variations can be distinguished via the dashed and plane box fillings. The yellow boxes show the pre-fit impact (top $x$ -axis) by varying each nuisance parameter by $\pm 1$ . The filled circles show the deviation of the fitted value for each nuisance parameter, $\hat{\theta}$ , from their nominal input value $\theta_0$ expressed in standard deviations with respect to their nominal uncertainties $\Delta\theta$ (bottom $x$ -axis). The error bars show the post-fit uncertainties on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show the fitted values and uncertainties of the normalization parameters that are freely floating in the fit. Pre-fit, these parameters have a value of one [190]. . . . .	162
7.11 Measured $VH$ simplified template cross sections times the $H \rightarrow b\bar{b}$ and $V \rightarrow \text{leptons}$ branching fractions in the medium and high $p_T^{V,t}$ bins [142]. . . . .	165

# List of tables

2.1	The fermions of the SM [18]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign. . . . .	15
2.2	The bosons of the SM [18]. The photon, weak bosons and gluons are gauge bosons arising from gauge symmetries, and carry the four fundamental forces of the SM. The recently discovered Higgs boson is the only fundamental scalar particle in the SM. . . . .	16
3.1	Overview of the different LHC runs [36, 37]. The average number of interactions per bunch-crossing is denoted as $\langle \mu \rangle$ (see Section 3.2.3), and is here averaged over the entire run. The luminosity is the peak instantaneous luminosity. Numbers for Run 3 are preliminary and are only provided to give an indication of expected conditions. . . . .	29
3.2	Quality selections applied to tracks, where $d_0$ is the transverse IP of the track, $z_0$ is the longitudinal IP with respect to the primary vertex and $\theta$ is the track polar angle (see Section 3.2.2 for the IP definitions). Silicon hits are hits on the pixel and SCT layers. Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [63]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. . . . .	47

5.1	Truth origins which are used to categorise the physical process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [63]. A truth-matching probability of less than 0.75 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The “OtherSecondary” origin includes tracks from photon conversions, $K_S^0$ and $\Lambda^0$ decays, and hadronic interactions. . . . .	76
5.2	Input features to the fake track classification NN. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [63], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split. “Primary vertex” (defined in Section 3.4.2) is abbreviated as PV. . . . .	78
5.3	Hyperparameters for the track classification model. . . . .	80
5.4	Good and fake track selection efficiencies for the combined $t\bar{t}$ and $Z'$ samples. Two working points are defined, cutting on the NN output at 0.06 and 0.12.	81
5.5	Cut values for the fake and $b$ -hadron decay track NNs for the two defined working points. Working point “B” cuts more aggressively on the NN outputs than WP “A”, removing more fake tracks but resulting in an increased loss of signal tracks (which here are all $b$ -hadron decay tracks). . . . .	82
6.1	Input features to the GN1 model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [63], while split hits are hits used on multiple tracks which have been identified as merged. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. The track leptonID is an additional input to the GN1Lep model [3]. “Primary vertex” (defined in Section 3.4.2) is abbreviated as PV. . . . .	91
6.2	A summary of GN1’s different classification networks used for the various training objectives, adapted from Ref. [3]. The hidden layers column contains a list specifying the number of neurons in each layer. ReLU activation is used through the network [122]. . . . .	96

---

6.3	The area under the ROC curves (AUC) for the track classification from GN1, compared to a standard multilayer perceptron (MLP) trained on a per-track basis. The unweighted mean AUC over the origin classes and weighted mean AUC (using as a weight the fraction of tracks from the given origin) is provided. GN1, which uses an architecture that allows track origins to be classified in a conditional manner as discussed in Section 6.4.3, outperforms the MLP model for both $t\bar{t}$ and $Z'$ jets. . . . .	115
7.1	Summary of the definitions of the different analysis regions . Signal enriched regions are marked with the label SR. There are regions with relatively large signal purity (HP SR) and with low purity (LP SR). Background enriched control regions are marked with the label CR. The shorthand “add” stands for additional small- $R$ jets, i.e. number of small- $R$ jets not matched to the Higgs-jet candidate. The medium and high $p_T^V$ regions are referred to as $Mp_T^V$ and $Hp_T^V$ , respectively [142]. . . . .	127
7.2	Selections applied to baseline and signal electrons and muons. Signal leptons are additionally required to satisfy a $p_T > 27 \text{ GeV}$ (25 GeV in the 1-lepton channel). . . . .	128
7.3	Event selection requirements for the boosted $VH$ , $H \rightarrow b\bar{b}$ analysis channels and sub-channels [142]. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region. The $\min[\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{small-}R \text{ jets})]$ selection is used to remove jets when the missing transverse momentum $\mathbf{E}_T^{\text{miss}}$ is pointing in the direction of the Higgs candidate, and the $\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T, \text{trk}})$ is used to reject events where the calorimeter missing transverse momentum $\mathbf{E}_T^{\text{miss}}$ is not pointing in the direction of the track-based missing transverse momentum $\mathbf{E}_{T, \text{trk}}^{\text{miss}}$ . The $\Delta y(V, H_{\text{cand}})$ quantifies the rapidity difference between the reconstructed vector boson and Higgs candidate ( $H_{\text{cand}}$ ). $m_{\ell\ell}$ refers to the invariant mass of the two leptons, while $p_T^{\ell_1}$ and $p_T^{\ell_2}$ refer to the $p_T$ of the first and second leptons, respectively. . . . .	130

7.4 Signal and background processes with the corresponding generators used for the nominal samples [142]. If not specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant ( $\alpha_s$ ). (★) The events were generated using the first PDF in the NNPDF3.0NLO set and subsequently reweighted to the PDF4LHC15NLO set [181] using the internal algorithm in POWHEG-Box v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for the $pp \rightarrow ZH$ process already includes the $gg \rightarrow ZH$ contribution. The $qq \rightarrow ZH$ process is normalised using the cross-section for the $pp \rightarrow ZH$ process, after subtracting the $gg \rightarrow ZH$ contribution. An additional scale factor is applied to the $qq \rightarrow VH$ processes as a function of the transverse momentum of the vector boson, to account for electroweak (EW) corrections at NLO. This makes use of the $VH$ differential cross-section computed with HAWK [182, 183]. . . . .	133
7.5 Different sources of uncertainty (i.e. variations in the model) considered for the $V+jets$ background, and the corresponding implementation. For each uncertainty, acceptance and shape uncertainties are derived. . . . .	138
7.6 0-lepton $W+jets$ nominal sample flavour composition and total event yield [190]. $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see Table 7.1). . . . .	139
7.7 1-lepton $W+jets$ nominal sample flavour composition and total event yield [190]. $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see Table 7.1). . . . .	140
7.8 0-lepton $Z+jets$ nominal sample flavour composition and total event yield [190]. $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see Table 7.1). . . . .	140
7.9 2-lepton $Z+jets$ nominal sample flavour composition and total event yield [190]. $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see Table 7.1). . . . .	140
7.10 $V+jets$ acceptance uncertainties. $W+jets$ SR and CR uncertainties marked with a superscript † are correlated. The 1L $W+jets$ medium-to-high $p_T^V$ uncertainty marked by * is applied as independent and uncorrelated NPs in both HP and LP signal regions [190]. . . . .	142
7.11 Diboson acceptance uncertainties. All uncertainties except channel extrapolation uncertainties are fully correlated between the $ZZ$ and $WZ$ processes and channels [190]. . . . .	146

---

7.12 Acceptance uncertainties for the $t\bar{t}$ background. The uncertainties on the 0- and 1-lepton channels are found to be similar, but are conservatively taken to be uncorrelated. The HP-to-LP SR uncertainty is applied in the HP SR. The Medium-to-high $p_T^V$ uncertainty is applied in the high $p_T^V$ bin, and is larger than the computed uncertainty in order to account for the known mismodelling of the $t\bar{t}$ $p_T$ spectrum. [190]. . . . .	148
7.13 Single-top acceptance uncertainties. Uncertainties are derived in the 1-lepton channel but applied in a correlated fashion to both the 0- and 1-lepton channels. The HP-to-LP SR uncertainty is applied in the HP SR. The Medium-to-high $p_T^V$ uncertainty is applied in the high $p_T^V$ bin. The channel relative acceptance uncertainty is applied in the 0-lepton channel. [190]. . .	149
7.14 Factors applied to the nominal normalisations of the $t\bar{t}$ , $W+\text{hf}$ , and $Z+\text{hf}$ backgrounds, as obtained from the likelihood fit [142]. The errors represent the combined statistical and systematic uncertainties. . . . .	155
7.15 Post-fit yields in the 0-lepton channel. Combined statistical and systematic uncertainties are shown [190]. . . . .	155
7.16 Post-fit yields in the 1-lepton channel. Combined statistical and systematic uncertainties are shown [190]. . . . .	156
7.17 Post-fit yields in the 2-lepton channel. Combined statistical and systematic uncertainties are shown [190]. . . . .	156
7.18 Breakdown of the absolute contributions to the uncertainty on the signal strength $\mu_{VH}^{bb}$ obtained from the (1+1)-POI fit. The average impact represents the average between the positive and negative uncertainties on $\mu_{VH}^{bb}$ . The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations [190]. . . . .	164