

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**

March 23, 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). A primary application of b -tagging is the study of the Higgs boson, which preferentially decays to a pair of b -quarks. A focus is placed on the high transverse momentum regime, which is a critical region in which to study the Standard Model, but also a region within which successful b -tagging becomes difficult.

As b -tagging relies on the successful 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 fb^{-1} of 13 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 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, 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^AT_EX 2 _{ε} using the “heptesis” class [1].

Contents

1	Introduction	12
2	Theoretical Framework	14
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
3	The Large Hadron Collider and the ATLAS Detector	28
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	42
3.3.4	The Trigger	43
3.4	Reconstructed Physics Objects	44
3.4.1	Tracks	44
3.4.2	Vertices	48
3.4.3	Jets	49
3.4.4	Leptons	51

3.4.5	Missing Transverse Momentum	53
3.5	Flavour Tagging Algorithms	53
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	57
4	Tracking and Flavour Tagging at High-p_T	59
4.1	Datasets	59
4.2	b -hadron Reconstruction Challenges	60
4.3	Investigations into High p_T b -hadron Tracking Improvements	64
4.3.1	Shared Hits	64
4.3.2	Global χ^2 Fitter Outlier Removal	64
4.4	Conclusion	67
5	Track Classification MVA	69
5.1	Machine Learning Background	69
5.1.1	Neural Networks	71
5.1.2	Training with Gradient Descent	73
5.2	Track Truth Origin Labelling	73
5.3	Fake Track Identification Tool	74
5.3.1	Model Inputs	76
5.3.2	Model Hyperparameters	78
5.3.3	Results	79
5.4	b -hadron Track Identification	80
5.5	Combined Approach	80
5.6	Conclusion	82
6	Graph Neural Network Flavour Tagger	84
6.1	Motivation	84
6.2	Graph Neural Network Theory	86
6.3	Experimental Setup	88
6.3.1	Datasets	88
6.4	Model Architecture	89
6.4.1	Model Inputs	89
6.4.2	Auxiliary Training Objectives	91

6.4.3	Architecture	92
6.4.4	Training	96
6.5	Results	97
6.5.1	<i>b</i> -tagging Performance	98
6.5.2	<i>c</i> -tagging Performance	100
6.5.3	Ablations	103
6.5.4	Inclusion of Low-Level Vertexing Algorithms	105
6.5.5	Jet Display Diagrams	105
6.5.6	Vertexing Performance	107
6.5.7	Track Classification Performance	113
6.6	Further Studies	115
6.6.1	Looser Track Selection	115
6.6.2	High Level Trigger	116
6.6.3	High Luminosity LHC	116
6.7	Conclusion	119
7	Boosted VHbb Analysis	123
7.1	Analysis Overview	124
7.1.1	Object Reconstruction	125
7.1.2	Event Selection	127
7.1.3	Control Regions	128
7.1.4	Background Composition	128
7.1.5	Data & Simulated Samples	131
7.1.6	Overview of Systematic Uncertainties	131
7.2	Experimental Uncertainties	133
7.3	Background Modelling Uncertainties	134
7.3.1	Truth Tagging	135
7.3.2	Sources of Systematic Modelling Uncertainties	135
7.3.3	Implementation of Variations	137
7.3.4	Vector Boson + Jets Modelling	138
7.3.5	Diboson Modelling	144
7.3.6	$t\bar{t}$ and single-top Modelling	147
7.3.7	Signal Modelling	148
7.4	Statistical Treatment	149
7.4.1	The Likelihood Function	149

7.4.2	Background Normalisations	152
7.4.3	Asimov Dataset & Expected Results	153
7.5	Results	153
7.5.1	Post-fit Results	154
7.5.2	Observed Signal Strength & Significance	158
7.5.3	Impact of Systematics	160
7.5.4	STXS Interpretation	162
7.6	Conclusion	165
8	Conclusion	166
8.1	Summary	166
8.2	Future Work	167
Bibliography		169
List of figures		184
List of tables		196

² Chapter 1

³ Introduction

- ⁴ This thesis describes efforts to improve the understanding of the Higgs boson and its
- ⁵ coupling to heavy flavour quarks, primarily through the development of algorithms
- ⁶ used to reconstruct and identify jets. The thesis is structured in the following manner:
- ⁷ Chapter 2 describes the theoretical foundations of the work presented in the rest of
- ⁸ the thesis.
- ⁹ Chapter 3 describes the ATLAS detector and the CERN accelerator complex. Details
- ¹⁰ of reconstructed physics objects and the b -jet identification algorithms are given.
- ¹¹ Chapter 4 provides an overview of the challenges facing successful charged particle
- ¹² trajectory (track) reconstruction and correspondingly b -jet identification, with a
- ¹³ particular focus on the high transverse momentum regime. Preliminary investigations
- ¹⁴ into reconstruction improvements are provided.
- ¹⁵ Chapter 5 describes the development of an algorithm which predicts the origins of
- ¹⁶ tracks. The tool is used to improve b -tagging performance by the identification and
- ¹⁷ removal of fake tracks before their input to the b -tagging algorithms.
- ¹⁸ Chapter 6 introduces a novel, monolithic b -jet identification algorithm which makes
- ¹⁹ use of graph neural networks and auxiliary training objectives.
- ²⁰ Chapter 7 describes the measurement of the associated production of a Higgs boson
- ²¹ decaying into a pair of b -quarks at high transverse momentum.
- ²² Chapter 8 contains some concluding remarks.

²³ The author's contribution to the work presented in this thesis is as follows.

²⁴ **Tracking:** The author was an active member of the Cluster and Tracking in Dense
²⁵ Environments group throughout their PhD, starting with their qualification task
²⁶ on the understanding of tracking performance at high transverse momentum. The
²⁷ author played a key role in software r22 validation studies for the tracking group,
²⁸ including the validation of the quasi-stable particle interaction simulation and the
²⁹ radiation damage Monte-Carlo simulation. The author helped design and improve
³⁰ several tracking software frameworks, and contributed to heavy flavour tracking
³¹ efficiency studies in dense environments. The author developed a tool to identify
³² and reject fake-tracks, which is being investigated for use in the upcoming tracking
³³ paper.

³⁴ ***b*-tagging:** The author has been an active member of the Flavour Tagging group
³⁵ since October 2020. The author played a key role in investigating the performance
³⁶ of the low level taggers at high transverse momentum and led studies into the
³⁷ labelling and classification of track origins. Based on work by Jonathan Shlomi [2],
³⁸ the author helped develop a new flavour tagging algorithm which offers a large
³⁹ performance improvement with respect to the current state of the art. The author
⁴⁰ was the primary editor of a public note associated with this work [3], which will
⁴¹ also be further developed in an upcoming paper. The author also contributed to
⁴² the proliferation of the new algorithm to the trigger, High Luminosity LHC, and
⁴³ $X \rightarrow bb$ use cases. The author also played a key role in software r22 validation
⁴⁴ studies for the Flavour Tagging group, including the validation of the quasi-stable
⁴⁵ particle interaction simulation. The author maintains and contributes to various
⁴⁶ software frameworks used in the Flavour Tagging group, including as lead developer
⁴⁷ of three packages, to create training datasets, pre-process samples for performance
⁴⁸ studies and a framework for training graph neural networks, and contributes to group
⁴⁹ documentation.

⁵⁰ **Higgs:** The author was an active member of the Boosted VHbb analysis group. The
⁵¹ author performed various studies deriving systematic uncertainties for the $V+jets$
⁵² and diboson backgrounds. The author also produced and maintained samples, ran fit
⁵³ studies and cross checks, and gave the diboson unblinding approval talk to the Higgs
⁵⁴ group. The author also contributed to the development of the analysis software.

55 **Chapter 2**

56 **Theoretical Framework**

57 The Standard Model (SM) of particle physics is the theory describing all known
58 elementary particles and their interactions via three of the four fundamental forces.

59 Developed by merging the successful theories of quantum mechanics and relativity
60 in the second half of the 20th century, the SM's position today at the centre of our
61 understanding of the nature of the Universe is firmly established by an unparalleled
62 level of agreement between the model's predictions and experimental results [4, 5].

63 The SM has predicted the discovery of the top and bottom quarks [6–8], the W
64 and Z bosons [9], and the tau neutrino [10]. The last missing piece of the SM to be
65 discovered was the Higgs boson, first theorised in the 1960s [11–13], and eventually
66 observed at the LHC in 2012 [14, 15]. After its discovery, much ongoing work has
67 been carried out performing detailed measurements of its mass and interactions with
68 other particles.

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

71 **2.1 The Standard Model**

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

75 Central to QFT is the Lagrangian density which describes the kinematics and
 76 dynamics of a field. Observations of conserved quantities are linked, via Noether's
 77 theorem, to symmetries which are expressed by the Lagrangian. Alongside Global
 78 Poincaré symmetry, the SM Lagrangian observes a local non-Abelian $SU(3)_C \otimes$
 79 $SU(2)_L \otimes U(1)_Y$ gauge symmetry. Gauge symmetries leave observable properties of
 80 the system unchanged when the corresponding gauge transformations are applied
 81 to the fields. The full Lagrangian of the SM can be broken up into distinct terms
 82 corresponding to the different sectors, as in Eq. (2.1). An overview of each sector is
 83 given in the following chapters.

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

84 The SM provides a mathematical description of how three of the four fundamental
 85 forces interact with the matter content of the Universe. The particle content of the
 86 SM consists of spin-1/2 fermions, listed in Table 2.1, and integral spin bosons listed
 in Table 2.2.

Generation	Leptons			Quarks		
	Flavour	Mass [MeV]	Charge [e]	Flavour	Mass [MeV]	Charge [e]
First	e	0.511	-1	u	2.16	$2/3$
	ν_e	$< 1.1 \times 10^{-6}$	0	d	4.67	$-1/3$
Second	μ	105.7	-1	c	1.27×10^3	$2/3$
	ν_μ	< 0.19	0	s	93.4	$-1/3$
Third	τ	1776.9	-1	t	173×10^3	$2/3$
	ν_τ	< 18.2	0	b	4.18×10^3	$-1/3$

Table 2.1: The fermions of the SM [16]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.

87

88 2.1.1 Quantum Electrodynamics

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

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

Table 2.2: The bosons of the SM [16]. 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 ψ . The field ψ 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 γ^μ (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 α 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_μ which transforms as $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu\alpha(x)$ must be added. The transformed interaction term

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

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

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

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

116 2.1.2 Quantum Chromodynamics

117 Quantum Chromodynamics (QCD) is the study of quarks, gluons and their interac-
118 tions. Quarks and gluons carry colour charge, which comes in three kinds, called
119 red, green and blue. While the $U(1)$ symmetry group in Section 2.1.1 was Abelian,
120 the QCD Lagrangian is specified by requiring invariance under transformations from
121 the non-Abelian $SU(3)$ group, making it a Yang-Mills theory [17] which requires the
122 addition of self-interacting gauge fields. The infinitesimal $SU(3)$ group generators
123 are given by $T_a = \lambda_a/2$, where λ_a are the eight Gell-Mann matrices. These span the
124 space of infinitesimal group transformations and do not commute with each other,
125 instead satisfying the commutation relation

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

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

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

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

131 where A_μ^a are the gluon fields and g_s is the strong coupling constant. The covariant
132 derivative is written as

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

133 The full QCD Lagrangian is then given by

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

134 Cubic and quartic terms of the gauge fields A_μ^a appear in the Lagrangian, leading to
135 the gluon's self interaction.

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

144 2.1.3 The Electroweak Sector

145 The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam
146 (GWS) model of electroweak interaction [18–20]. The Lagrangian is specified by
147 requiring invariance under the symmetry group $SU(2)_L \otimes U(1)_Y$, as motivated by a
148 large amount of experimental data. Here, $SU(2)_L$ is referred to as weak isospin and
149 $U(1)_Y$ as weak hypercharge.

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

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

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

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

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

160 The weak force violates parity conservation [21–23], i.e. invariance under parity
161 transformations (mirror reflections). Only left handed fermions participate in the
162 weak interaction. Since there is no other force through which neutrinos interact with
163 other particles, there are no right handed neutrinos in the Standard Model.

164 2.2 The Higgs Mechanism

165 The Brout-Englert-Higgs mechanism (henceforth just the “Higgs mechanism”) is the
166 process through which the fundamental particles of the SM acquire mass [11–13].
167 Experimentally it was known that the weak force had a low effective strength, which
168 was suggestive of a massive mediating gauge particle. However, directly adding
169 mass to the weak gauge bosons violates the non-Abelian symmetry of the SM.
170 Instead, the gauge bosons gain mass through the interaction with a scalar Higgs
171 field which results from the spontaneous breakdown of symmetry as discussed in
172 Section 2.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as discussed
173 in Section 2.2.2. Section 2.2.3 described some basic phenomenology of the Higgs
174 particle relevant to hadron colliders.

¹⁷⁵ 2.2.1 Electroweak Symmetry Breaking

¹⁷⁶ Spontaneous symmetry breaking (SSB) is a key part of the Higgs mechanism. It
¹⁷⁷ is the transition of a physical system from a state of manifest symmetry to a state
¹⁷⁸ of hidden, or *broken*, symmetry. In particular, this applies to physical systems
¹⁷⁹ where the Lagrangian observes some symmetry, but the lowest energy vacuum states
¹⁸⁰ do not exhibit that same symmetry. In other words, the symmetry is broken for
¹⁸¹ perturbations around the vacuum state.

¹⁸² Consider the case in which the gauge fields from the local $SU(2)_L \otimes U(1)_Y$ symmetry
¹⁸³ group (discussed in Section 2.1.3) are coupled to a complex scalar field $\phi = \phi(x)$,
¹⁸⁴ transforms as a weak isospin doublet. Omitting the kinetic term of the gauge fields,
¹⁸⁵ and writing $\phi^2 \equiv \phi^\dagger \phi$, the Lagrangian is

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

¹⁸⁶ where the covariant derivative is given by

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

¹⁸⁷ and T^a are the generators of $SU(2)$. The potential term $V(\phi)$ is made up of a
¹⁸⁸ quadratic and quartic term in the scalar field ϕ , which each contain an arbitrary
¹⁸⁹ parameter, respectively λ and μ . The quartic term gives the field self-interaction, and
¹⁹⁰ cannot be negative as this would lead to a potential that was unbounded from below.
¹⁹¹ The quadratic term can be positive or negative. In the case where the quadratic term
¹⁹² is positive, it is interpreted as a mass term for the scalar field. By choosing $\mu^2 < 0$
¹⁹³ the field becomes unphysical due to its negative mass. The shape of the potential in
¹⁹⁴ this case is shown in Fig. 2.1. Note that in the case of the Standard Model, the scalar
¹⁹⁵ field ϕ is a complex doublet, and so the corresponding potential is 5-dimensional.
¹⁹⁶ In order to obtain a physical interpretation of the Lagrangian in Eq. (2.14) for the
¹⁹⁷ case where $\mu^2 < 0$, the field ϕ is expanded around the vacuum state. The vacuum
¹⁹⁸ expectation value (VEV) is the expected value of the field ϕ which minimises the
¹⁹⁹ potential $V(\phi)$ (equivalently the expected value of the field operator ϕ when the
²⁰⁰ 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

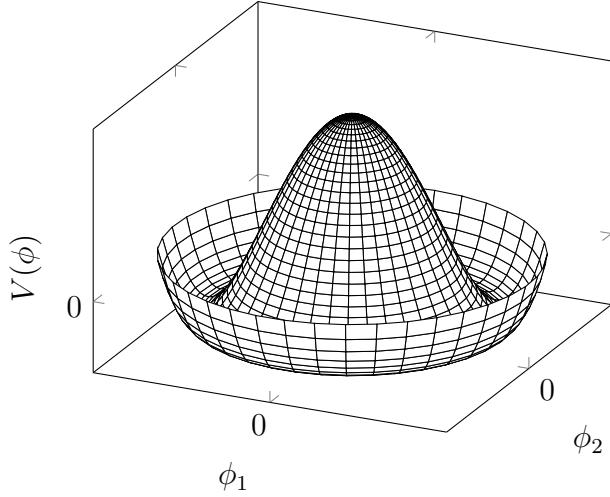


Figure 2.1: The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin.

201 gives a VEV of

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

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

209 Adding the particle content back to the theory by expanding the field around the
210 vacuum state, and making a transformation to the unitary gauge to remove unphysical
211 Nambu-Goldstone modes (which arise in the context of global symmetries [24, 25]),
212 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_μ^\pm , Z_μ and A_μ in terms of the original fields A_μ^a and B_μ . 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 θ_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 ϕ 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 λ 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.

237 Spontaneous symmetry breaking has modified the original symmetry group of the SM
238 $SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \otimes U(1)_Q$. Three broken generators from the symmetry
239 group $SU(2)_L \times U(1)_Y$ have been absorbed into the definition of the physical weak
240 vector bosons, giving them mass. The same methodology can be used to generate
241 the fermion masses, as shown in the next section.

242 2.2.2 Fermionic Yukawa Coupling

243 Adding the masses of the fermions by hand breaks the gauge invariance of the
244 theory. Instead, we can use a Yukawa coupling between the fermion fields and the
245 Higgs field in order to generate mass terms after spontaneous electroweak symmetry
246 breakdown [19]. In this way, the fermion masses are determined by both the respective
247 couplings to the Higgs field and the VEV of the Higgs field itself, which sets the
248 basic mass scale of the theory.

249 The Higgs field ϕ transforms as an $SU(2)$ doublet with $Y = 1/2$, as does the left-
250 handed fermion field ψ_L . The right-handed fermion field ψ_R transforms as an $SU(2)$
251 singlet.

252 Charged Lepton Masses

253 The renormalisable and gauge invariant coupling between a fermionic field ψ and a
254 scalar Higgs field ϕ can be written as

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

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

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

259 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 ϕ provides
260 the first term in Eq. (2.22) which is quadratic in the electron field, and can therefore

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

264 **Quark Masses**

265 The down-type quarks acquire their mass analogous to the leptons, with $\psi_L = (u_L, d_L)$
 266 and $\psi_R = d_R$ for the first quark generation. Mass is generated for the up-type quarks
 267 using the conjugate field to ϕ which transforms under $SU(2)$ as a doublet with
 268 $Y = -1/2$. The conjugate field $\tilde{\phi}$ is constructed as

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

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

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

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

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

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

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

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

283 2.2.3 Higgs Sector Phenomenology

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

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

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

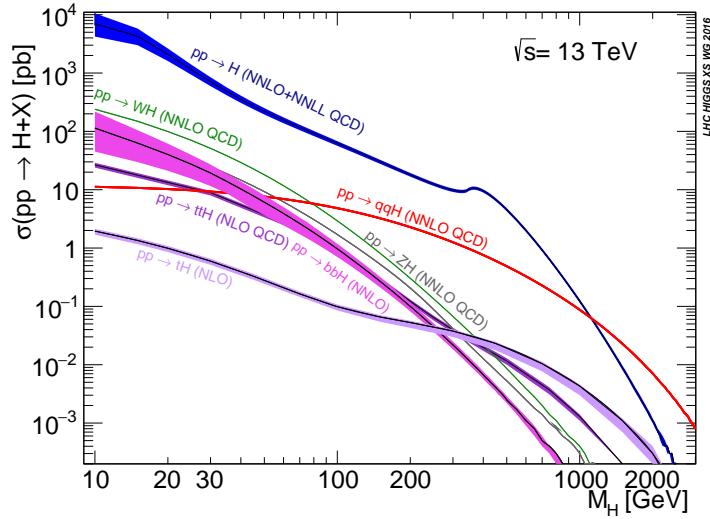


Figure 2.2: Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13 \text{ TeV}$ [26]. Uncertainties are shown in the shaded bands. At $m_H = 125 \text{ GeV}$, Higgs boson production is dominated by gluon-gluon fusion, vector boson fusion, associated production with vector bosons, and top quark fusion.

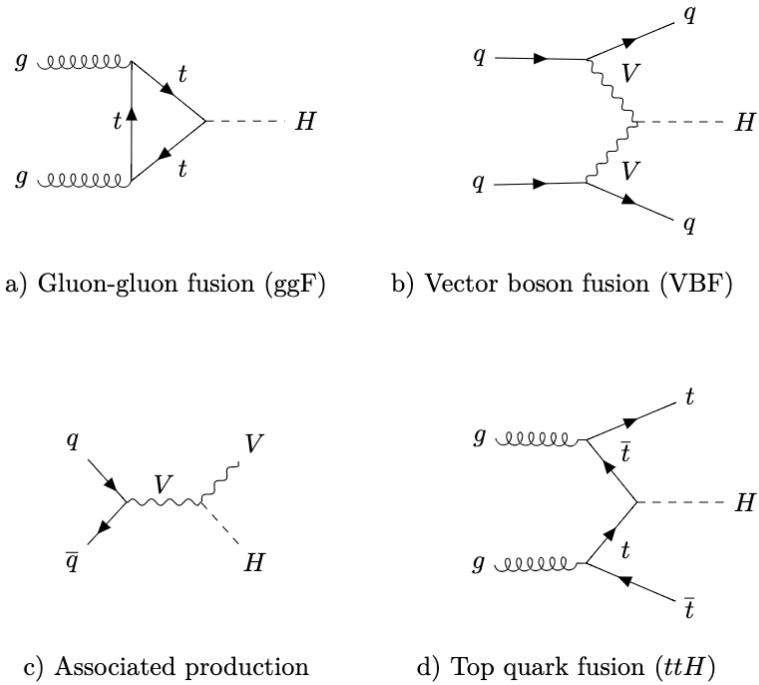


Figure 2.3: Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125 \text{ GeV}$ at a centre of mass energy $\sqrt{s} = 13 \text{ TeV}$.

³¹³ modes (around 0.2% for $H \rightarrow \gamma\gamma$), these decay channels were instrumental in the
³¹⁴ initial discovery of the Higgs due to the low level of background processes which
³¹⁵ mimic the final state [14, 15].

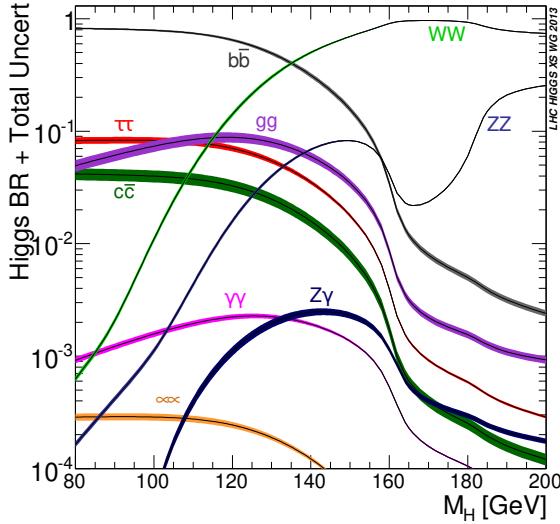


Figure 2.4: Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [26]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, the Higgs predominantly decays to a pair of b -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of W bosons.

³¹⁶ This thesis presents a measurement of the Higgs bosons production rate using events
³¹⁷ with a Higgs boson produced in association with vector boson and decaying to a pair
³¹⁸ of b -quarks, i.e. $pp \rightarrow VH(b\bar{b})$. The $H \rightarrow b\bar{b}$ decay mode directly probes the Higgs
³¹⁹ coupling to fermions, and more specifically to the bottom quark. This coupling was
³²⁰ first observed in 2018 [27, 28]. Ongoing work measuring the coupling strengths, in
³²¹ particular in the high energy regime, is the focus of the analysis presented in this
³²² thesis in Chapter 7.

323 Chapter 3

324 The Large Hadron Collider and the
325 ATLAS Detector

326 Since the completion of its construction in 2008, the Large Hadron Collider (LHC) [29]
327 at CERN has extended the frontiers of particle physics through significant increases
328 in centre of mass energy and luminosity compared with previous collider experiments.
329 The LHC accelerates bunches of protons around a 27 km ring until they are travelling
330 just 3 m s^{-1} slower than the speed of light, at which point they are made to collide.
331 The proton bunches travel round the ring 11,000 times per second in two concentric
332 beams, which are guided by superconducting magnets cooled using liquid helium to
333 -271.3°C (1.9 K). The beams travel in opposite directions around the ring and are
334 crossed at four locations so that collisions between protons can take place. Around
335 these collision points four specialised detectors, ALICE [30], CMS [31], LHCb [32] and
336 ATLAS [33], are located to capture information about the products of the collisions.

337 In this chapter, a brief overview of the LHC and the accelerator complex at CERN
338 is given in Section 3.1. The coordinate system used at the ATLAS detector and
339 other common definitions are introduced in Section 3.2. In Section 3.3 the different
340 ATLAS detector systems are described, and in Section 3.4 information about various
341 commonly used reconstructed objects is given. Finally, Section 3.5 provides an
342 introduction to flavour tagging at ATLAS.

³⁴³ 3.1 The Large Hadron Collider

³⁴⁴ The LHC is operated in multi-year *runs* during which beams of protons are circulated
³⁴⁵ and collided. Between runs there are periods of shutdown while the accelerator and
³⁴⁶ detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC
³⁴⁷ collided proton bunches, each containing more than 10^{11} particles, 20 million times
³⁴⁸ per second, providing 7 TeV proton-proton collisions at instantaneous luminosities
³⁴⁹ of up to $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The centre-of-mass energy was increased to 8 TeV
³⁵⁰ in 2012. Over the course of Run 1, 26.4 fb^{-1} of usable integrated luminosity was
³⁵¹ recorded. Run 2, which spanned 2015–2018, further increased the proton-proton
³⁵² collision energy to 13 TeV. During Run 2 the bunch spacing was reduced, leading
³⁵³ to a collision rate of 40 MHz. Over the course of Run 2 a total usable integrated
³⁵⁴ luminosity of 139 fb^{-1} was recorded. 2022 marked the beginning of Run 3 which,
³⁵⁵ with a higher center of mass energy and peak luminosity, is expected to culminate in
³⁵⁶ an approximate tripling of the dataset size. A summary of key information about
³⁵⁷ 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	8×10^{33}
Run 2	2015–2018	13	34	25	$1\text{--}2 \times 10^{34}$
Run 3	2022–2025	13.6	50	25	2×10^{34}

Table 3.1: Overview of the different LHC runs [34,35]. 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 performance.

³⁵⁸ An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is
³⁵⁹ at the final stage of a chain of accelerators which incrementally step-up the energy
³⁶⁰ of incoming protons. The first accelerator is Linac2 (which has been replaced by
³⁶¹ Linac4 in 2020), a linear accelerator which accelerates negative hydrogen ions to an
³⁶² energy of 160 MeV. Upon leaving Linac4, the ions are stripped of both electrons
³⁶³ and the resulting protons are fed into the Proton Synchrotron Booster (PSB), which
³⁶⁴ increases the energy of the protons to 2 GeV. The protons leaving the PSB are passed
³⁶⁵ to the Proton Synchrotron (PS), which increases the energy to 26 GeV, and then
³⁶⁶ from the PS to the Super Proton Synchrotron (SPS) which further increases the

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

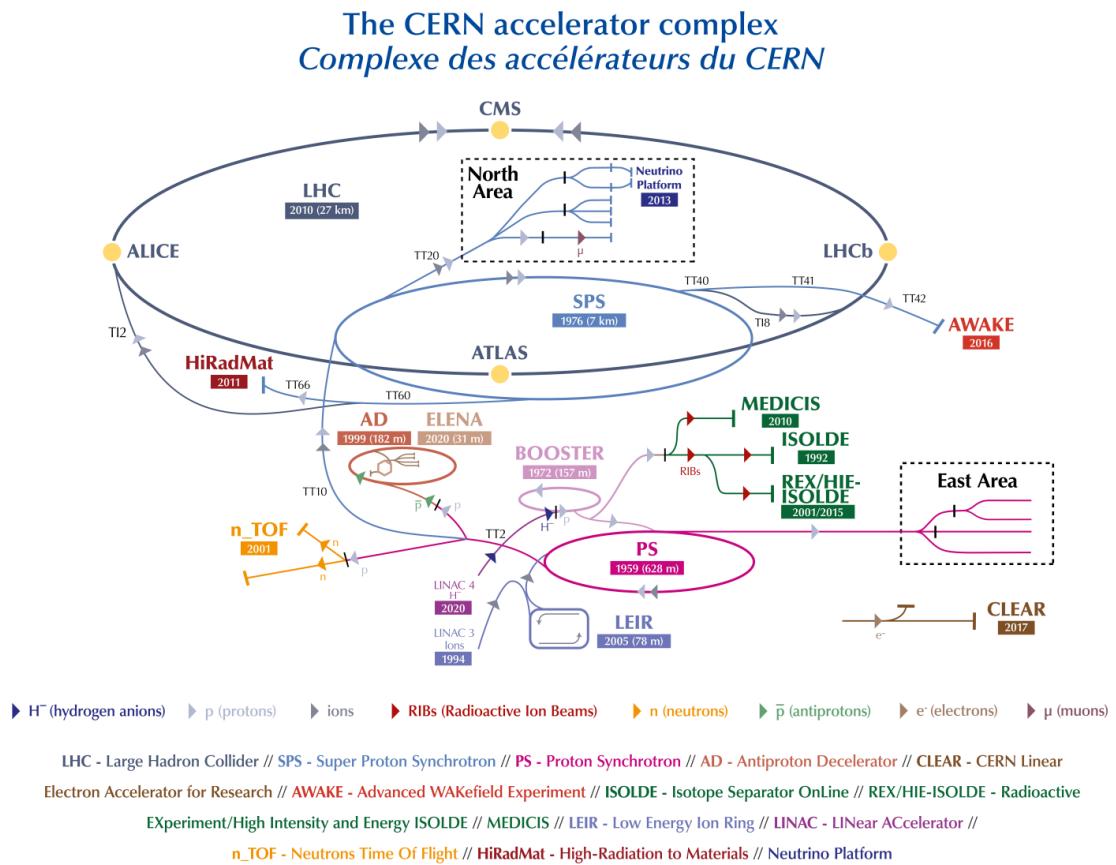


Figure 3.1: An overview of the CERN accelerator complex [36]. The LHC is fed by a series of accelerators starting with Linac2 (or Linac4 from 2020). Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

369 3.2 Coordinate System & Collider Definitions

³⁷⁰ 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
³⁷¹ Section 3.2.2, and definitions for some frequently occurring concepts and quantities
³⁷² is provided in Section 3.2.3.

³⁷⁴ 3.2.1 ATLAS Coordinate System

³⁷⁵ The origin of the coordinate system used by ATLAS is the nominal interaction point
³⁷⁶ in the centre of the detector. As shown in Fig. 3.2, the z -axis points along the
³⁷⁷ direction of the beam pipe, while the x -axis points from the interaction point to the
³⁷⁸ centre of the LHC ring, and the y -axis points upwards. The transverse plane lies
³⁷⁹ in x - y while the longitudinal plane lies along the z -axis. A cylindrical coordinate
³⁸⁰ system with coordinates (r, ϕ) is used in the transverse plane, where r is the radius
³⁸¹ from the origin and ϕ is the azimuthal angle around the z -axis.

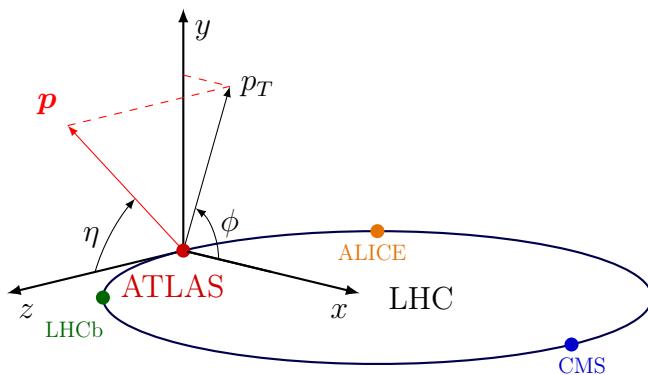


Figure 3.2: The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [37].

³⁸² The polar angle θ is commonly specified in terms of the pseudorapidity η , defined as

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

³⁸³ The pseudorapidity is a convenient quantity to work with as differences in η are
³⁸⁴ invariant under Lorentz boosts.

³⁸⁵ The transverse momentum p_T of an object is the sum in quadrature of the momenta
³⁸⁶ in the transverse plane

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

387 Angular distance between two objects is measured in units of ΔR and is defined as

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

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

390 3.2.2 Track Parameterisation

391 The trajectories of charged particle tracks are parameterised as a helix which is fully
392 specified using five parameters: $(d_0, z_0, \phi, \theta, q/p)$. The transverse and longitudinal
393 impact parameters (IP) d_0 and z_0 specify the closest approach of the trajectory of a
394 particle to an given origin, where the hard scatter primary vertex (see Section 3.4.2)
395 is used in this thesis. ϕ and θ are the azimuthal and polar angles respectively, and
396 q/p is the measured charge on the track¹ divided by the scalar 3-momentum. Fig. 3.3
397 shows each of these parameters diagrammatically.

398 Impact parameter significances are defined as the IP divided by its corresponding
399 uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. When used in flavour tagging
400 (see Chapter 4), track IP significances are lifetime signed according to the track's
401 direction with respect to the jet axis and the primary vertex [39]. The signed IP
402 significances is positive if the track crosses the jet axis in front of the primary vertex
403 and negative if the crossing is behind the primary vertex.

404 3.2.3 Hadron Collider Definitions

405 Cross Section

406 The cross section σ is closely related to the probability of an interaction between
407 two colliding particles, and is analogous to an effective cross-sectional area of the
408 particles. The cross section of a process depends on the transition matrix element
409 and a phase space integral. At hadron colliders such as the LHC, the proton-proton

¹Reconstructed charged particles are assumed to have a charge of ± 1 .

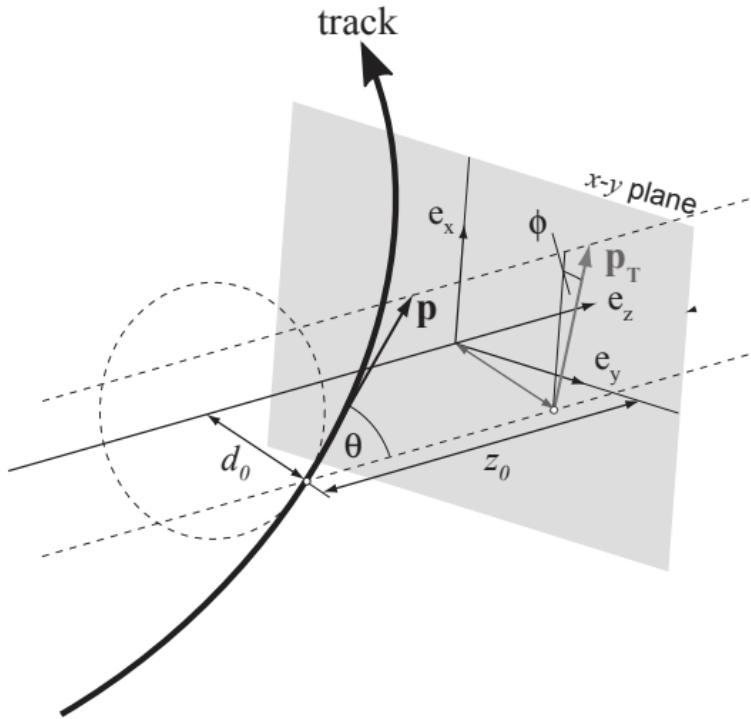


Figure 3.3: The track parameterisation used at the ATLAS detector. Five coordinates ($d_0, z_0, \phi, \theta, q/p$) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum \mathbf{p} and 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. [38].

⁴¹⁰ cross section can be factorised as

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

⁴¹¹ The partonic cross section can be calculated at high energies such as those found at
⁴¹² the LHC, while the parton distribution functions (PDFs) have to be extracted from
⁴¹³ experimental results.

⁴¹⁴ Luminosity

⁴¹⁵ The total number of proton-proton collisions N is related to the total pp cross σ
⁴¹⁶ section by the integrated luminosity L , as in

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

⁴¹⁷ The instantaneous luminosity \mathcal{L} relates the cross section to the number of collisions
⁴¹⁸ per unit time. For two colliding bunched proton beams, it is defined as

$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} = \frac{fn_1 n_2}{4\pi \sigma_x \sigma_y}, \quad (3.6)$$

⁴¹⁹ where n_1 and n_2 are the number of protons in the colliding bunches, f is the bunch
⁴²⁰ crossing frequency, and σ_x and σ_y are the RMS width of the beam in the horizontal
⁴²¹ and vertical directions.

⁴²² The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In
⁴²³ total, 139 fb^{-1} of usable physics data was collected over the three-year run. The
⁴²⁴ uncertainty on the total integrated luminosity is 1.7% [40].

⁴²⁵ Pile-up

⁴²⁶ At the centre of the ATLAS detector, bunches of more than 10^{11} protons are collided.
⁴²⁷ Each bunch-crossing is called an *event*. There is generally one hard proton-proton
⁴²⁸ scatter per event. Additional interactions are typically relatively soft ($low-p_T$) and
⁴²⁹ are known as *pile-up*. Pile-up from interactions within the same bunch-crossing is
⁴³⁰ known as *in-time* pile-up while residual signatures from previous bunch-crossings
⁴³¹ is known as *out-of-time* pile-up. The number of pile-up interactions is denoted μ ,

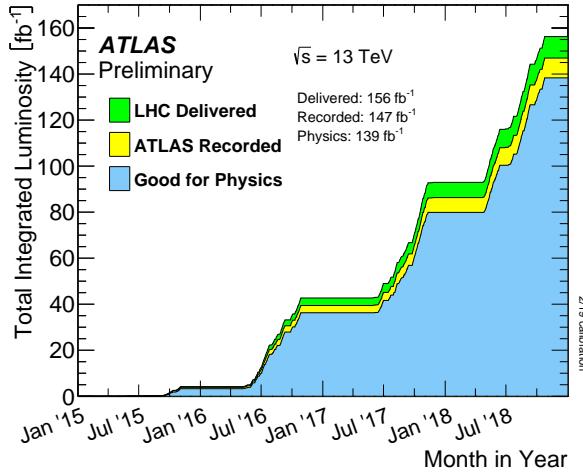


Figure 3.4: Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [35]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.

432 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.

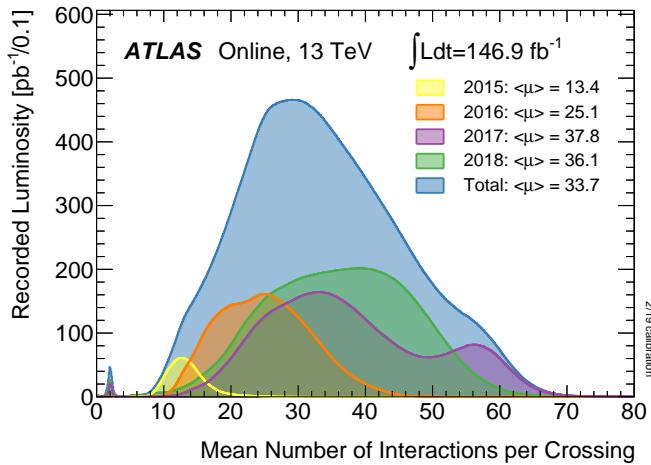


Figure 3.5: Average pile-up profiles measured by ATLAS during Run 2 [35].

⁴³⁴ 3.3 The ATLAS Detector

⁴³⁵ The ATLAS² detector is made up of several specialised sub-detectors which are
⁴³⁶ arranged concentrically around the nominal interaction point at the centre of the
⁴³⁷ detector, as shown in Fig. 3.6. The detector is designed to cover nearly the entire solid
⁴³⁸ angle around the collision point. In this section a brief overview of each sub-detector
⁴³⁹ is given, in order of increasing radial distance from the point of collision. The inner
⁴⁴⁰ tracking detector is described in Section 3.3.1, the electromagnetic and hadronic
⁴⁴¹ calorimeters in Section 3.3.2, the muon spectrometer in Section 3.3.3, and finally the
⁴⁴² trigger is described in Section 3.3.4. More complete information on the detector can
⁴⁴³ be found in Ref. [33], while an overview of physics performance is given in [41].

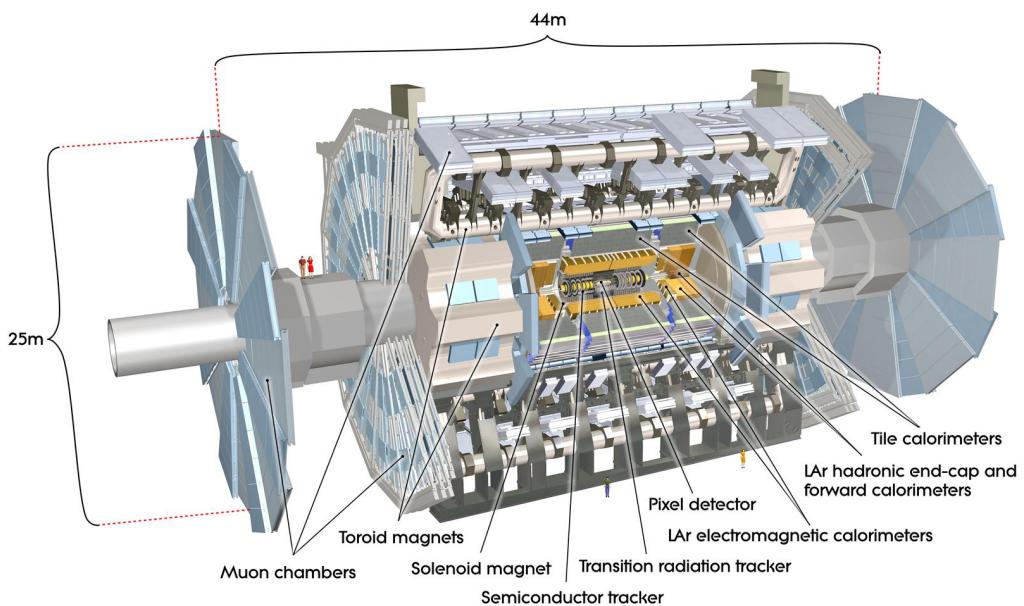


Figure 3.6: A 3D model of the entire ATLAS detector [42]. Cutouts to the centre of the detector reveal the different sub-detectors which are arranged in concentric layers around the nominal interaction point.

²A Toroidal LHC ApparatuS.

⁴⁴⁴ 3.3.1 Inner Detector

⁴⁴⁵ The inner-detector system (ID) provides high-resolution charged particle trajectory
⁴⁴⁶ tracking in the range $|\eta| < 2.5$. The ID is immersed in a 2 T axial magnetic field,
⁴⁴⁷ produced by a superconducting solenoidal magnet, which enables the measurement
⁴⁴⁸ of particle momentum and charge. After Run 3, the ID will be replaced by the
⁴⁴⁹ ITk [43, 44].

⁴⁵⁰ The inner detector is made up of several sub-systems, shown in Figs. 3.7 and 3.8.
⁴⁵¹ The high-granularity silicon pixel detector covers the innermost region and typically
⁴⁵² provides four space-point measurements per track. It is followed by the silicon
⁴⁵³ microstrip tracker (SCT), which usually provides a further four space-point measure-
⁴⁵⁴ ments (8 hits) per track. These silicon detectors are complemented by the Transition
⁴⁵⁵ Radiation Tracker (TRT), which enables radially extended track reconstruction up
⁴⁵⁶ to $|\eta| = 2.0$ and typically provides 33 (38) additional hits in the barrel (end-cap).

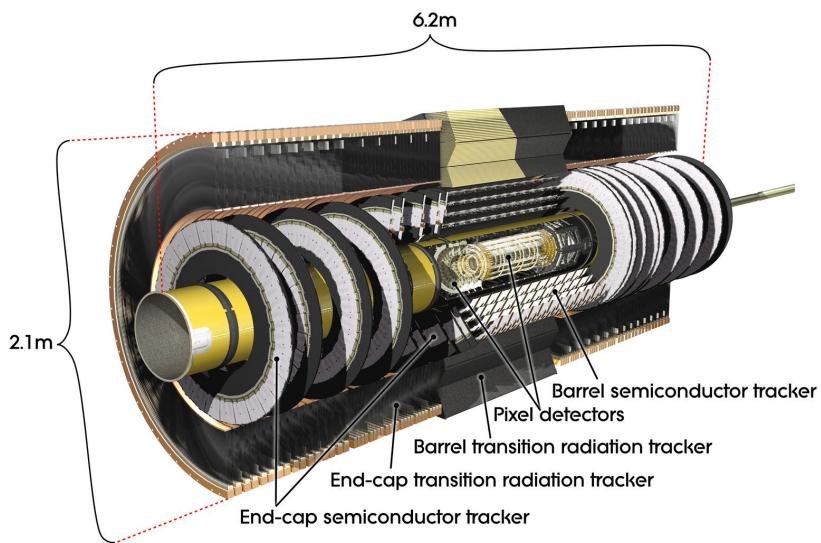


Figure 3.7: A 3D model of the ATLAS ID showing the pixel, SCT and TRT sub-detectors [45].

⁴⁵⁷ The target inverse momentum resolution for the combined ID measurement is
⁴⁵⁸ parameterised as a function of the track transverse momentum and polar angle [41].

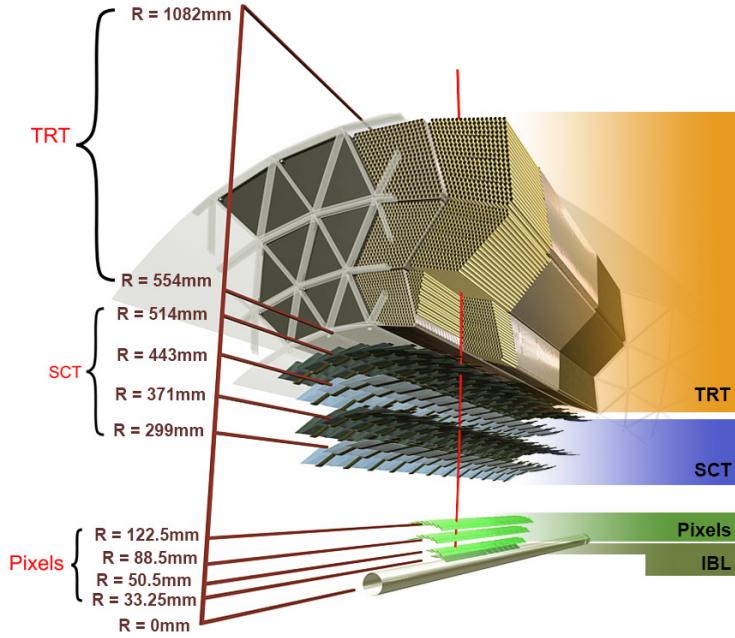


Figure 3.8: A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [38].

⁴⁵⁹ The parameterisation is given by

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

⁴⁶⁰ where \oplus denotes a sum in quadrature. For low- p_T tracks (e.g. $p_T \approx 500 \text{ MeV}$) in the
⁴⁶¹ central region this corresponds to a relative error of approximately 0.01%. Meanwhile
⁴⁶² for high- p_T tracks (e.g. $p_T \approx 100 \text{ GeV}$) in the central region this corresponds to
⁴⁶³ a relative error of approximately 4%. The momentum resolution generally good
⁴⁶⁴ enough to correctly identify the sign of the charge on particles up to the highest
⁴⁶⁵ energies expected at the LHC. The transverse impact parameter resolution $\sigma(d_0)$ is
⁴⁶⁶ parameterised similarly as

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

⁴⁶⁷ Typical uncertainties for the transverse IP resolution are 230 μm and 11 μm for low
⁴⁶⁸ and high- p_T tracks in the central region, respectively.

⁴⁶⁹ Pixel Detector

⁴⁷⁰ The silicon pixel detector is comprised of four cylindrical barrels at increasing radii
⁴⁷¹ from the beamline, and four disks on each side. The innermost barrel layer is the
⁴⁷² insertable B-layer (IBL), shown in Fig. 3.9. The IBL was installed before Run 2 [46,47]
⁴⁷³ and lies approximately just 33 mm from the beam axis. The second-to-innermost layer
⁴⁷⁴ is often referred to as the B-layer. The specification of the pixel detector determines
⁴⁷⁵ the impact parameter resolution and the ability to reconstruct primary and secondary
⁴⁷⁶ vertices. The detector is required to have a high granularity (i.e. resolution) to
⁴⁷⁷ maintain the low occupancy required to resolve nearby particles. Individual pixels
⁴⁷⁸ are 50 μm in the transverse direction $R\phi$ and 400 μm in the longitudinal z direction
⁴⁷⁹ (250 μm for the IBL). Cluster positions have a resolution of approximately 10 μm in
⁴⁸⁰ $R\phi$ and 100 μm in z .

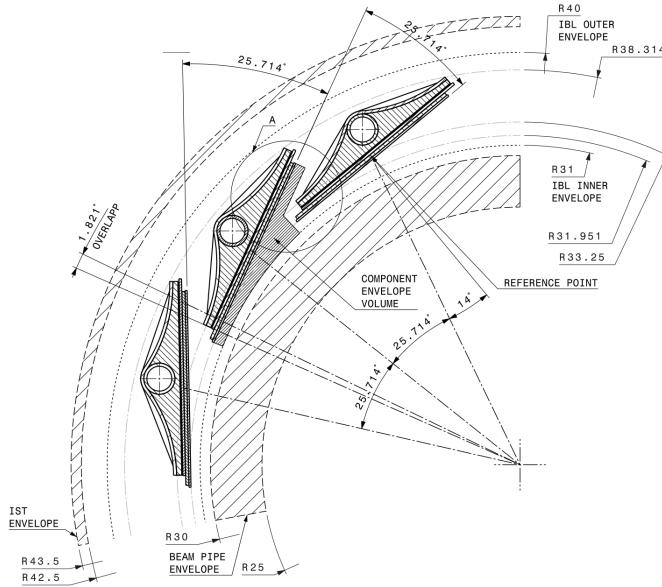


Figure 3.9: A schematic cross-sectional view of the ATLAS IBL [46].

⁴⁸¹ Semi-Conductor Tracker (SCT)

⁴⁸² The SCT is made up of four concentric barrel layers in the central region, and nine
⁴⁸³ disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,
⁴⁸⁴ with a small stereo angle (20 mrad) between the two layers enabling the z -coordinate

485 to be measured from a pair of strip measurements. The SCT typically provides
486 four precision space-point measurements (eight strip measurements) per track in
487 the barrel region. These have intrinsic uncertainties of $17\text{ }\mu\text{m}$ in the transverse
488 direction $R\phi$, and $580\text{ }\mu\text{m}$ in the longitudinal direction z [48]. The measurements
489 provide a contribution to the measurement of charged particle momentum and impact
490 parameter.

491 **Transition Radiation Tracker (TRT)**

492 The TRT is a straw-tube tracker which complements the higher-resolution silicon-
493 based tracks by offering a larger number of hits per track (typically more than 30)
494 and a long lever arm, which aids the accurate measurement of particle momentum.
495 It is made up of approximately 300 000 drift tubes with a diameter of 4 mm which
496 are filled with an argon/xenon gas mixture. The walls of each tube are electrically
497 charged, and a thin conducting wire runs along the center. When a charged particle
498 traverses a tube, it ionises the gas and the resulting liberated electrons drift along
499 the electric field to the wire, where an associated charge is registered. In the barrel
500 the straws run parallel to the z -axis and therefore the TRT only provides tracking
501 information in $R\phi$. Straws are arranged radially in the end-caps. The resulting
502 two-dimensional space-points have a resolution of approximately $120\text{ }\mu\text{m}$. The spaces
503 between the straws are filled with a polymer which encourages the emission of
504 transition radiation, aiding electron identification.

505 **3.3.2 Calorimeters**

506 The calorimeter system measures the energy of incident particles over the range
507 $|\eta| < 4.9$. There are two main sub-systems: the electromagnetic calorimeter (ECal),
508 which focuses on the measurement of electrons and photons, and the hadronic
509 calorimeter (HCal), which measures the energy of hadrons. A schematic view of the
510 calorimeter system is shown in Fig. 3.10. Upon entering the calorimeter, incident
511 particles will interact with the detector material to produce a shower of secondary
512 particles with reduced energies. The charge deposited in this process is measured to
513 reconstruct the energy of the initial incident particle. The two calorimeter sub-systems

514 must provide strong containment of showering particles to prevent punch-through of
515 EM and hadronic particles to the HCal and muon systems respectively.

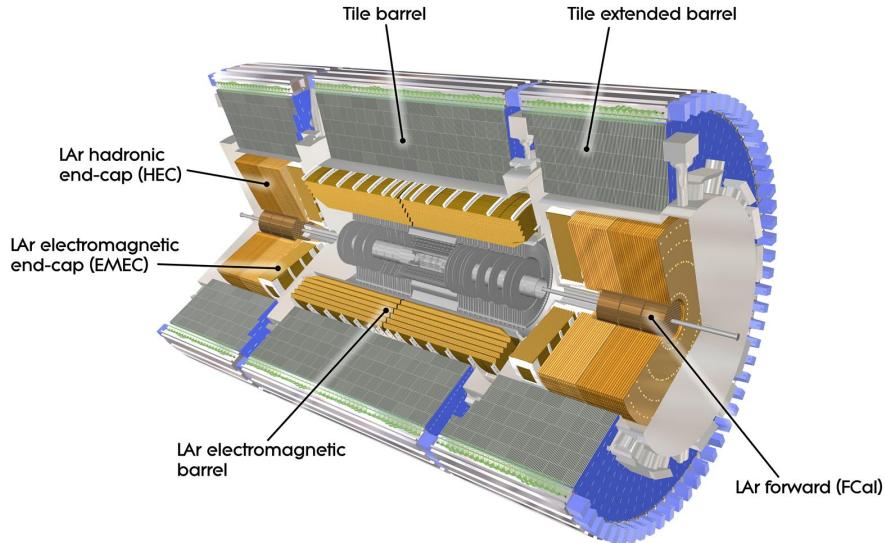


Figure 3.10: The ATLAS calorimeters [49]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic end-caps.

516 Liquid Argon (LAr) Electromagnetic Calorimeter

517 The more granular lead/liquid-argon ECal covers the region $|\eta| < 3.2$ and is split
518 into barrel (covering $|\eta| < 1.475$) and end-cap (covering $1.375 < |\eta| < 3.2$) regions.
519 EM calorimetry works by encouraging electrons and photons to interact with electrici-
520 cally charged particles in detector material via bremsstrahlung ($e \rightarrow e\gamma$) and pair
521 production ($\gamma \rightarrow e^+e^-$). The EM calorimeter uses lead absorber plates to initiate
522 EM showers, resulting in secondary particles which ionise the surrounding liquid
523 argon. The charge is collected on copper electrodes and read out. The accordion
524 geometry of the ECal allows for a full coverage in ϕ without any azimuthal cracks.
525 The energy resolution of the LAr calorimeter is made up of a sampling and a constant
526 term, which are summed in quadrature to produce the overall energy resolution. The
527 sampling term contributes approximately $10\%/\sqrt{E}$, while the constant term adds an
528 additional 0.7%. Photons with moderate transverse energy $E_T \approx 50$ GeV have an

529 overall energy resolution of at most 1.6% over most of the pseudorapidity range. At
 530 lower $E_T \approx 10 \text{ GeV}$, the resolution is degraded to approximately 5%. The resolution
 531 measurements are obtained from test beam data [41].

532 **Hadronic Tile Calorimeter**

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

539 The hadronic energy resolution of the HCal is parameterised as a function of the
 540 hadron's transverse energy

$$\sigma(E_T)/E_T = 50\%/\sqrt{E_T} \oplus 3\%, \quad (3.9)$$

541 corresponding to a energy resolution of 11% (6.5%) for a hadron with E_T of approxi-
 542 mately 10 GeV (50 GeV) [50].

543 **3.3.3 Muon Spectrometer**

544 Due to their higher mass, muons easily pass unimpeded through the ID and calorime-
 545 ters and therefore require specialised detectors for their measurement. The Muon
 546 Spectrometer (MS) is made up of dedicated tracking and triggering hardware, as
 547 shown in Fig. 3.11. The precision tracking system uses three layers of monitored drift
 548 tubes with a barrel region covering $|\eta| < 1.2$ and end-caps covering $1 < |\eta| < 2.7$.
 549 The inner layers of the end-caps use cathode strip chambers to better cope with the
 550 high occupancy in the forward region. The trigger system is comprised of resistive
 551 plate chambers in the barrel region covering $|\eta| < 1.0$ and thin gap chambers in
 552 the end-cap regions covering $1 < |\eta| < 2.4$. A set of three superconducting air-core
 553 toroidal magnets, each made up of eight coils, is used in each of the barrel and
 554 end-caps to deflect the muons as they pass through the MS, allowing their momentum
 555 and charge to be measured from the direction and magnitude of curvature. The

556 toroidal magnets generate a field which is largely orthogonal to the muon trajectories
 557 which allows for maximum deflection. The transverse momentum resolution (mea-
 558 sured for combined ID and muon tracks, see Section 3.4.4) has been measured to be
 559 approximately 1.7% in the central region for muons from J/ψ decays, increasing to
 2.3% for muons in the forward regions [51].

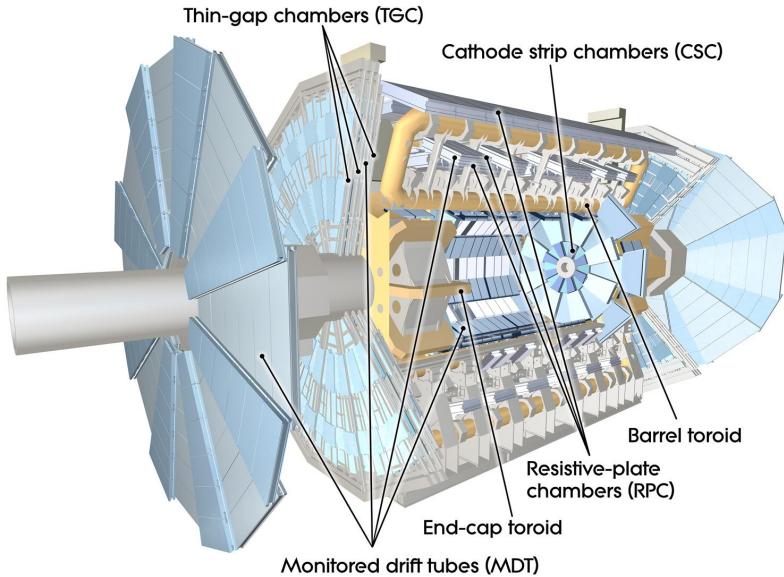


Figure 3.11: The ATLAS muon spectrometer [52].

560

561 3.3.4 The Trigger

562 The 25 ns bunch spacing used over the course of Run 2 corresponds to a bunch-
 563 crossing or event rate of 40 MHz (see Table 3.1). If the full information for the
 564 detector was written out for each event, this would correspond to the generation
 565 of 60 TB of data each second. This is more than can be feasibly read out from the
 566 hardware, processed and stored, requiring the use of a trigger system which quickly
 567 makes a decision about whether or not an event is potentially interesting and should
 568 be kept for further analysis. The trigger system is comprised of two levels which
 569 aim to identify various signatures, such as electrons, muons, taus, photons, and jets
 570 (including b -jets), as well as events with large total or missing transverse energy. The
 571 hardware-based Level-1 (L1) trigger uses coarse information from the calorimeters
 572 and MS to accept events at an average rate of 100 kHz approximately 2.5 μ s after the

573 event. After the L1 trigger, the software-based High Level Trigger (HLT) makes use
574 of 40 000 CPU cores to make a final selection on surviving events in approximately a
575 few hundred milliseconds. The final event read-out rate is approximately 1.2 kHz,
576 corresponding to 1.2 GB s^{-1} of permanent data storage. More information is provided
577 in [53].

578 3.4 Reconstructed Physics Objects

579 Event reconstruction is the process of analysing the output from the detector to
580 determine the type and properties of particles present in an event. The reconstructed
581 event provides information about the underlying physics process that led to these
582 observable final state particles. Events passing the trigger selection (described in
583 Section 3.3.4) undergo offline reconstruction, which makes use of the full information
584 from the detector. Reconstruction and analysis of events relies on the extensive
585 ATLAS software stack, see Ref. [54] for more information.

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

588 3.4.1 Tracks

589 The reconstructed trajectories of charged particles are referred to as *tracks*. Tracks
590 are reconstructed from the energy depositions (called *hits*) left by the particles as
591 they traverse the inner detector. Tracks are used in the reconstruction of other
592 objects, including vertices and jets, so their accurate reconstruction is a critical
593 task. A comprehensive introduction to ATLAS tracking is available in Ref. [55],
594 while specific optimisations for dense environments are detailed in Refs. [56, 57]. An
595 overview of track reconstruction is given below.

596 Space-point Formation (Clustering)

597 When a charged particle traverses a silicon layer, charge can be collected in more
598 than one pixel or strip. This is due to the incident angle of the particles with respect
599 to the sensor, and also the drift of electrons between sensors caused by the magnetic

600 field. Clusters (also called *hits* or *space-points*) are formed by clustering neighbouring
601 pixels or strips and estimating locations of space-points using the shape and energy
602 distribution of the clusters.

603 **Track Finding**

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

620 **Ambiguity Solving**

621 Ambiguity solving was introduced as part of the ATLAS New Tracking effort [55],
622 which was intended to improve track reconstruction performance in dense envi-
623 ronments. In the ambiguity solver, track candidates are processed individually in
624 descending order of a track score. The track score quantifies the likelihood of the
625 track corresponding to the trajectory of a real particle. Scoring uses a number of
626 variables, including the number and positions of hits (preferring hits in more precise
627 regions of the detector), the transverse momentum of the track and the track fit
628 quality. The track fit quality describes the quality of the track as the χ^2 divided
629 by the number degrees of freedom on the track. A preference for high transverse

630 momentum tracks promotes the successful reconstruction of the more physically
 631 interesting energetic particles, and suppresses the large number of wrong hits assigned
 632 to low momentum tracks. The ambiguity solver also penalises tracks with missing
 633 hits on the innermost detector layers.

634 During the processing of a track candidate, the track is cleaned (whereby problematic
 635 hits are removed), and, if the resulting track satisfies the quality selection criteria, a
 636 high precision fit of the track parameters using the surviving hits is performed. The
 637 high precision fit makes full use of all available information, and uses an updated
 638 position and uncertainty estimate for each cluster obtained from a Neural Network
 639 (NN) [58]. If the track has reached this stage without being rejected by passing various
 640 quality requirements, it is re-scored and returned to the list of track candidates. If
 641 the same track is then processed again without requiring modification, it is added to
 642 the final track collection. Track candidates that fall below certain quality threshold
 643 are rejected. This selection does allow for the possibility of a track having small
 644 number of shared hits, as detailed in Table 3.2.

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	≥ 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 PV and θ 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 [57]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

645 Neural Network Cluster Splitting

646 As part of track cleaning, shared hits are classified by a NN to determine if they are
 647 compatible with the characteristic features of a merged cluster [56, 58]. A merged
 648 cluster is one made up of a combination of energy deposits from more than one
 649 particle, which have become merged due to the closeness of the associated particles
 650 and the limited resolution of the detector. It is common for clusters to become
 651 merged in dense environments, as discussed in Section 4.2. If the cluster is predicted
 652 to be merged it is labelled as being freely shareable, or *split*. Hits not compatible
 653 with the merged hypothesis can still be shared by a limited number of tracks, but
 654 come with a penalty for the track which may hinder its acceptance into the final
 track collection.

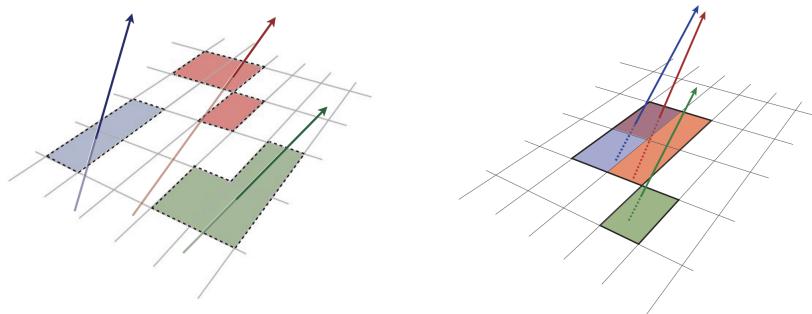


Figure 3.12: Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [57].

655

656 Pseudotracking

657 Pseudotracking uses Monte Carlo truth information to group together all the hits
 658 left by each truth particle. Each collection of hits which, as a unit, satisfies basic
 659 quality requirements is directly used in a full resolution track fit. If the track fit is
 660 successful, a “pseudotrack” track is created and stored. If the track fit fails, or the
 661 collection of hits does not pass the basic quality requirements (for example because
 662 of a lack of hits) then the particle is said to be un-reconstructable. In this way,
 663 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. [59] 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 [56].

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* (PV). Primary vertices are iteratively reconstructed with tracks using the iterative vertex finder (IVF) [60]. The *hard scatter vertex* of an event is chosen as the primary vertex whose associated tracks have the largest sum of transverse momentum squared, $\Sigma(p_T^2)$.

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 [61, 62]. 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 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

692 that secondary vertices from b -hadron decays are reconstructed separately to tertiary
 693 vertices from $b \rightarrow c$ decay chains.

694 3.4.3 Jets

695 Jets are an aggregate reconstructed object corresponding to a collection of collimated
 696 stable particles which results from the presence of a quark or gluon. Jets are built
 697 by clustering constituent objects (e.g. tracks or calorimeter clusters) using a jet
 698 finding algorithm, for example the anti- k_t algorithm [63], which is implemented in
 699 FASTJET [64].

700 Objects can be associated to jets in one of two ways. The first is via a geometrical
 701 matching in ΔR (see) The second is via a ghost association [65], where the object is
 702 assigned a negligible momentum and re-clustered into the jet after its formation.

703 Jets from pile-up interactions are suppressed using the Jet Vertex Tagger (JVT)
 704 algorithm, which uses the transverse momenta of tracks to identify jets from pile-up
 705 interactions [66].

706 EMTopo Jets

707 EMTopo jets are reconstructed from noise-suppressed topological clusters (topoclus-
 708 ters) of calorimeter energy depositions [67]. The clustering uses the energy significance
 709 of each cell, defined as

$$S_{\text{cell}} = \frac{E_{\text{cell}}}{\sigma_{\text{noise, cell}}}, \quad (3.10)$$

710 where E_{cell} is the energy measured in a given calorimeter cell, and $\sigma_{\text{noise, cell}}$ is the
 711 expected level of noise on the cell (e.g. from pile-up interactions). Topoclusters are
 712 formed from a seed cell with a large S_{cell} , and expanded by iteratively adding neigh-
 713 bouring cells with a sufficiently large energy significance. Collections of topoclusters
 714 are then clustered into a jet using the anti- k_t algorithm with a radius parameter of
 715 0.4 (small- R jets) or 1.0 (large- R jets). More information, including information on
 716 the calibration of the topocluster jet energy scale, is available in Ref. [67].

717 Particle Flow Jets

718 Particle-flow (PFlow) jets are reconstructed from particle-flow objects [68] using
719 the anti- k_t algorithm with a radius parameter of 0.4. Particle-flow objects integrate
720 information from both the ID and the calorimeters, improving the energy resolution
721 at high transverse momenta and reducing pile-up contamination. The PFlow jet
722 energy scale is calibrated according to Ref. [69].

723 Large- R Jets

724 Large- R jets have a radius parameter $R = 1.0$ and are built by clustering topological
725 calorimeter clusters using the anti- k_t algorithm [70]. The large radius parameter
726 is especially useful for containing the decay products of a boosted Higgs boson, as
727 discussed in Chapter 7. Due to their large size, large- R jets benefit from a grooming
728 procedure called trimming which remove soft contaminants inside the jet [71, 72].
729 Trimming aims to remove jet constituents from pile-up and the underlying event,
730 which helps to improve the jet mass resolution and its robustness to varying levels
731 of pile-up. The jet mass is computed using a combination of information from the
732 calorimeters and ID, and a calibration to data is applied [73].

733 Track-jets

734 Track-jets are built by clustering tracks using the anti- k_t clustering algorithm. They
735 are associated to large- R jets as sub-jets and used to identify large- R jets containing
736 b -hadrons. The radius parameter is allowed to vary with transverse momentum such
737 that a broader cone (up to $R = 0.4$) is used for low- p_T track-jets and a narrower
738 cone (down to $R = 0.02$) for high- p_T track-jets [74, 75]. The narrower cone is better
739 suited to clustering highly collimated jet constituents at high- p_T .

740 Jet Flavour Labels

741 Jet flavour labels are assigned to small- R jets according to the presence of a truth
742 hadron within $\Delta R(\text{hadron}, \text{jet}) < 0.3$ of the jet axis. If a b -hadron is found the jet is
743 labelled a b -jet. In the absence of a b -hadron, if a c -hadron is found the jet is called

⁷⁴⁴ a c -jet. If no b - or c -hadrons are found, but a τ is found in the jet, it is labelled as a
⁷⁴⁵ τ -jet, else it is labelled as a light-jet.

⁷⁴⁶ PFlow jets are used to train the algorithms discussed in Chapter 5 and Chapter 6.

⁷⁴⁷ Jet Track Association

⁷⁴⁸ Tracks are associated to small- R jets using a ΔR association cone, the width of which
⁷⁴⁹ decreases as a function of jet p_T , with a maximum cone size of $\Delta R \approx 0.45$ for jets
⁷⁵⁰ with $p_T = 20$ GeV and minimum cone size of $\Delta R \approx 0.25$ for jets with $p_T > 200$ GeV.
⁷⁵¹ If a track is within the association cones of more than one jet, it is assigned to the
⁷⁵² jet which has a smaller $\Delta R(\text{track}, \text{jet})$.

⁷⁵³ 3.4.4 Leptons

⁷⁵⁴ Electrons and muons leave characteristic signatures that are picked up in the ECal
⁷⁵⁵ and MS respectively. The reconstruction of both types of charged lepton is briefly
⁷⁵⁶ outlined below.

⁷⁵⁷ Electrons

⁷⁵⁸ A diagrammatic view of electron reconstruction is shown in Fig. 3.13. Electrons
⁷⁵⁹ candidates are reconstructed by matching PV-compatible³ inner detector tracks
⁷⁶⁰ to topological calorimeter clusters. The track-cluster matching criteria takes into
⁷⁶¹ account the significant energy loss of the electron due to bremsstrahlung. If a match
⁷⁶² is found, a refit of the track is performed using the Gaussian Sum Filter (GSF) [76],
⁷⁶³ which better handles trajectory reconstruction in the presence of bremsstrahlung.
⁷⁶⁴ Various identification criteria are then applied to the candidates using a likelihood-
⁷⁶⁵ based (LH) method to improve purity. These include requirements on the track
⁷⁶⁶ quality and cluster matching, the shape of electromagnetic shower in the ECal,
⁷⁶⁷ leakage into the HCal, and the amount of transition radiation detected in the TRT.
⁷⁶⁸ Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may
⁷⁶⁹ also be applied. A full description can be obtained from Ref. [77].

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

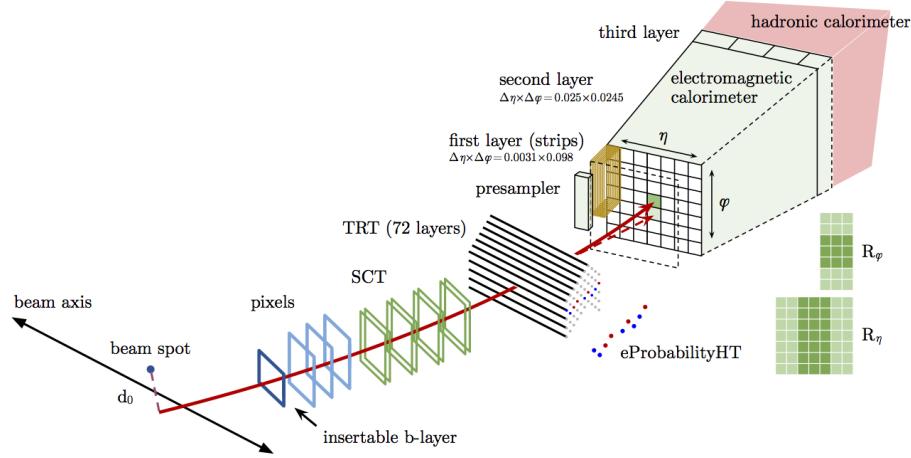


Figure 3.13: A sketch of electron reconstruction using the ATLAS detector [77]. Electron reconstruction makes use of the entire ID and the calorimeters.

770 Muons

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

781 After reconstruction, candidate undergo an identification processes which helps to
 782 efficiently identify prompt muons whilst rejecting background signals (e.g. non-
 783 prompt muons from pion and kaon decays, the punch-through of a hadron from the
 784 calorimeter, or the semi-leptonic decay of a heavy flavour hadron). Combined muon
 785 identification takes into account discrepancies in the p_T and charge measurements
 786 in the MS and ID, and the χ^2 of the combined track fit. Selections on the number
 787 of hits in the ID and MS are also applied. At the medium identification working
 788 point, approximately 96% of prompt muons with $20 \text{ GeV} < p_T < 100 \text{ GeV}$ are

789 successfully identified. On top of the identification requirements, a number of
790 isolation requirements can also be applied to further suppress background signals.

791 More information on muon reconstruction, identification and isolation can be found
792 in Ref. [51].

793 3.4.5 Missing Transverse Momentum

794 An imbalance in the final state transverse momentum can occur as a result of
795 incomplete measurement of the final state particles. In particular, neutrinos are
796 not measured by the detector and contribute to the missing transverse momentum
797 $\mathbf{E}_T^{\text{miss}}$. Incomplete detector acceptance and inaccuracies in the reconstruction of the
798 final state can also contribute to the missing transverse momentum of an event. In
799 order to calculate the missing transverse momentum, the negative vector sum of
800 the momentum of all photons, leptons and small- R jets with $p_T > 20 \text{ GeV}$ is taken.
801 The momenta of tracks associated to the primary vertex are also taken into account.
802 The magnitude of $\mathbf{E}_T^{\text{miss}}$ is written E_T^{miss} . More information about missing transverse
803 momentum reconstruction is provided in [78].

804 3.5 Flavour Tagging Algorithms

805 Many ATLAS analyses rely on flavour tagging, which is the identification of jets
806 containing heavy flavour (HF) hadrons (b -hadrons and c -hadrons) as opposed to those
807 containing only light-flavour hadrons or gluons (collectively referred to as light-jets).
808 In particular, b -tagging is the identification of jets originating only from b -hadrons
809 (i.e. b -jets). Flavour tagging is a critical component of the physics programme of the
810 ATLAS experiment. It is of crucial importance for the study of the Standard Model
811 (SM) Higgs boson and the top quark, which decay preferentially to b -quarks [79, 80],
812 and additionally for several Beyond the Standard Model (BSM) resonances that
813 readily decay to heavy flavour quarks [81].

814 In this section some common definitions used in flavour tagging are given in Sec-
815 tion 3.5.1, followed by a description of the characteristic properties of b -jets in

816 Section 3.5.2. The various b -jet identification algorithms (also called *taggers*) used in
817 ATLAS in Sections 3.5.3 and 3.5.4.

818 3.5.1 Common Definitions

819 Each flavour tagging algorithm outputs a discriminant value that can be used to
820 select jets of the signal flavour. The efficiency of a flavour tagging algorithm is
821 defined as the fraction of signal jets which are correctly identified by the tagger.
822 Given N^{signal} signal jets, $N_{\text{tagger}}^{\text{signal}}$ of which have passed the algorithm's selection and
823 $N_{\text{untagger}}^{\text{signal}}$ of which have not, the efficiency is given by

$$\varepsilon = \frac{N_{\text{tagged}}^{\text{signal}}}{N_{\text{tagged}}^{\text{signal}} + N_{\text{untagged}}^{\text{signal}}}. \quad (3.11)$$

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

$$\text{fake rate} = \frac{N_{\text{tagged}}^{\text{background}}}{N_{\text{tagged}}^{\text{background}} + N_{\text{untagged}}^{\text{background}}}. \quad (3.12)$$

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

$$\text{rejection power} = \frac{1}{\text{fake rate}}. \quad (3.13)$$

828 A *fixed-cut working point* (WP) defines the corresponding selection applied to the
829 tagging discriminant in order to achieve a given efficiency on the inclusive $t\bar{t}$ sample.

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. [82] 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 $\tau \approx 1.5$ ps [83]. This lifetime corresponds to a proper decay length $c\tau \approx 450$ μm . In the rest frame of the detector, the typical *b*-hadron travels a distance

$$d = \gamma\beta c\tau \approx \gamma c\tau \quad (3.14)$$

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 (the distance varies due to the interleaved structure). 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.

b-hadrons decay weakly to on average four or five collimated stable particles [84]. 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.

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

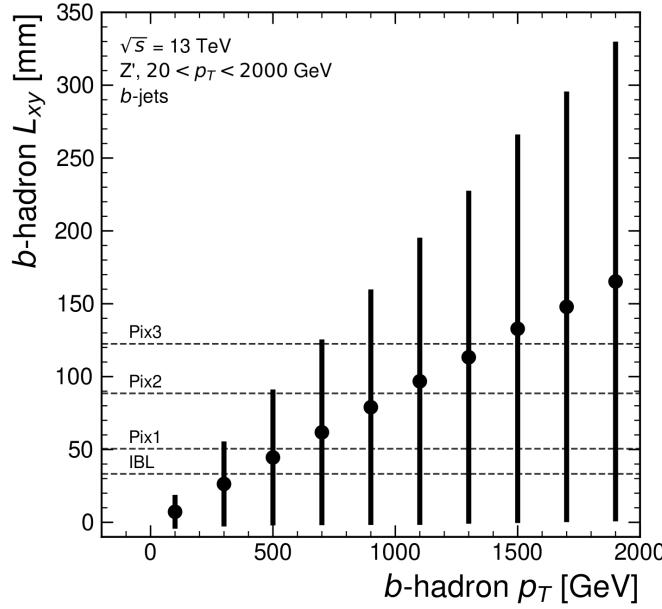


Figure 3.14: The truth b -hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed b -jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.

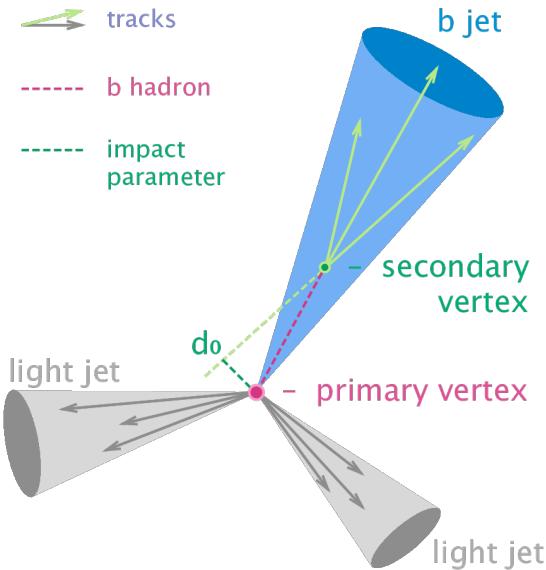


Figure 3.15: Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [85]. The b -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter d_0 is a characteristic property of the trajectories of b -hadron decay products.

860 (as shown in Fig. 3.15). Furthermore, since it is common for the b -hadron to decay
861 to a c -hadron with non-negligible lifetime, tertiary vertices can be found within
862 b -jets resulting from $b \rightarrow c$ decay chains. Additionally, as the b - or c -hadron decays
863 semileptonically in approximately 40% of cases [16], the presence of a reconstructed
864 electron or muon inside a jet gives information about the jet flavour.

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

869 3.5.3 Low-level Algorithms

870 The low-level algorithms are the first step in the b -tagging process. They take inputs
871 information about the jet and its associated tracks and reconstruct distinct features
872 of the experimental signature of heavy flavour jets. The low-level algorithms are a
873 combination of manually optimised reconstruction algorithms, for example SV1 and
874 JetFitter which reconstruct displaced decay vertices (see Section 3.4.2) and IPxD
875 which discriminates based on track IPs, and trained machine learning models such
876 as RNNIP and DIPS which use IP and hit information from a variable number of
877 tracks [86, 87].

878 3.5.4 High-level Algorithms

879 The high-level algorithms combine the outputs of the independently optimised low-
880 level algorithms using a multivariate approach to produce a discriminant value for
881 each jet. For example the MV2c10 algorithm, used in the analysis described in
882 Chapter 7, consists of a boosted decision tree which combines the outputs of IPxD,
883 SV1 and JetFitter [61, 88, 89]. The working point is tuned to achieve an average b -jet
884 efficiency of 70% on simulated $t\bar{t}$ events. At this efficiency working point, rejection
885 factors for c -jets and light-jets are approximately 9 and 304 respectively.

886 The current ATLAS flavour tagger, DL1r [89], is part of a series of taggers that
887 use a deep neural network to combine the outputs of the low-level algorithms. The

888 low-level algorithms used by the different high-level taggers DL1, DL1r and DL1d is
889 shown in Fig. 3.16.

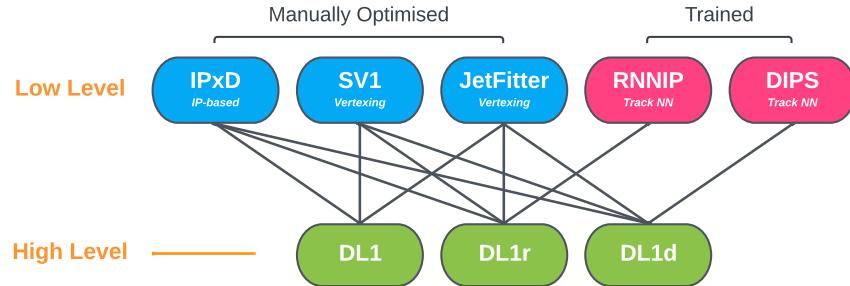


Figure 3.16: An overview of different high and low level taggers used in ATLAS. The low level taggers are IPxD, SV1 and JetFitter, and RNNIP and DIPS [61, 62, 86, 87]. The outputs of these taggers are fed into the high-level taggers DL1, DL1r and DL1d [89, 90].

890 **Chapter 4**

891 **Tracking and Flavour Tagging at**
892 **High- p_T**

893 The various flavour tagging algorithms (introduced in Section 3.5) work by identi-
894 fying the unique signatures of heavy flavour (HF) jets (b - and c -jets). Ultimately,
895 the tagging algorithms use input information about the reconstructed jet and its
896 associated tracks. Successful b -tagging therefore relies on the efficient and accurate
897 track reconstruction, especially for tracks corresponding to the products of HF decays.
898 In this chapter studies investigating the challenges facing track reconstruction and
899 flavour tagging at high- p_T are discussed.

900 The chapter is structured as follows. In Section 4.1 an introduction to the datasets
901 used in this thesis is given. A summary of the challenges facing tracking and b -
902 tagging at high transverse momentum is provided in Section 4.2. Some preliminary
903 investigations into improving tracking in the high- p_T regime are investigated in
904 Section 4.3. In Section 4.4 the conclusions of the chapter are given.

905 **4.1 Datasets**

906 This thesis makes extensive use of two simulated datasets which are described in this
907 section. The datasets are made up of simulated SM $t\bar{t}$ and BSM Z' events initiated by
908 proton-proton collisions at a center of mass energy $\sqrt{s} = 13$ TeV. The Z' sample is
909 constructed in such a manner that it has a relatively flat jet p_T spectrum up to 5 TeV
910 and decays democratically to equal numbers of b -, c - and light-jets. The generation

of the simulated event samples includes the effect of multiple pp interactions per bunch crossing with an average pile-up of $\langle \mu \rangle = 40$, which includes the effect on the detector response due to interactions from bunch crossings before or after the one containing the hard interaction.

The $t\bar{t}$ events are generated using the POWHEGBox v2 generator [91–94] at next-to-leading order with the NNPDF3.0NLO [95] set of parton distribution functions (PDFs). The h_{damp} parameter¹ is set to 1.5 times the mass of the top-quark (m_{top}) [96], with $m_{\mathrm{top}} = 172.5$ GeV. The events are interfaced to PYTHIA 8.230 [97] to model the parton shower, hadronisation, and underlying event, with parameters set according to the A14 tune [98] and using the NNPDF2.3LO set of PDFs [99]. Z' events are generated with PYTHIA 8.2.12 with the same tune and PDF set. The decays of b - and c -hadrons are performed by EVTGEN v1.6.0 [100]. Particles are passed through the ATLAS detector simulation [101] based on GEANT4 [102].

Jets are required to have a pseudorapidity $|\eta| < 2.5$ and $p_{\mathrm{T}} > 20$ GeV. Additionally, 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 [66].

4.2 b -hadron Reconstruction Challenges

This section some of the difficulties in the accurate reconstruction of the b -hadron decay at high- p_{T} associated reconstruction difficulties in Section 4.2. 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 ($p_{\mathrm{T}} > 250$ GeV) this task becomes difficult due to a combination of effects, which are 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.

¹The h_{damp} parameter 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.

940 This is compounded by the increased decay length of b -hadrons (and c -hadrons) at
 941 high- p_T , which means that the decay products have less of an opportunity to diverge
 942 before reaching the first tracking layers of the detector (shown in Fig. 4.1). If the
 943 weak decay of the b -hadron takes place close enough to a detector layer, or if the
 944 particles are otherwise sufficiently collimated, charge deposits left by nearby particles
 945 may not be resolved individually, instead being reconstructed as merged clusters [57].

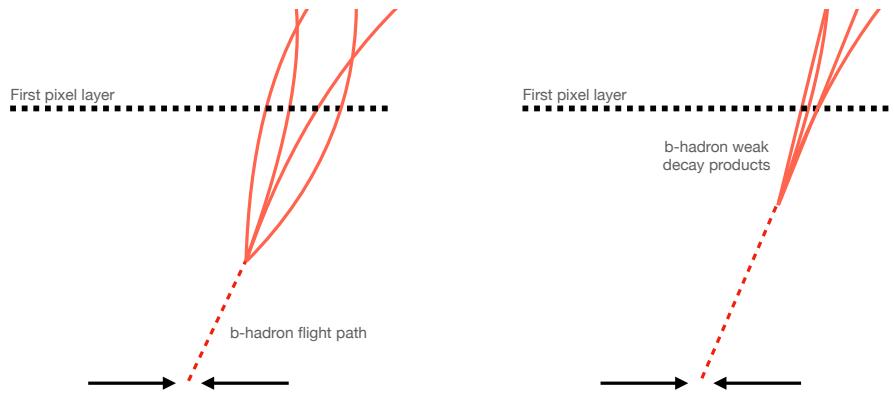


Figure 4.1: At lower p_T (left) the decay length of the b -hadron is on average reduced, and the decay tracks are less collimated. At higher p_T (right) the b -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching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

946 As discussed in Section 3.4.1, merged clusters are generally rare, and so shared
 947 hits generally predict bad tracks and are correspondingly penalised during track
 948 reconstruction. However, in the core of high p_T b -jets the density of particles is high
 949 enough that the probability of cluster merging increases dramatically. Successful
 950 reconstruction of such tracks requires the presence of shared hits to be effectively
 951 dealt with, but in the standard reconstruction the presence of these can instead
 952 impair the successful reconstruction of the track. Furthermore, HF decays may also
 953 take place inside the tracking detectors themselves, which at best leads to missing
 954 measurements on the most sensitive detector layers, and at worst can lead to wrong
 955 inner layer hits being added to displaced tracks, since the reconstruction process
 956 penalises tracks without inner layer hits.

957 The above effects create two distinct but related problems for b -tagging. The first is
 958 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 $p_T > 2 \text{ TeV}$.

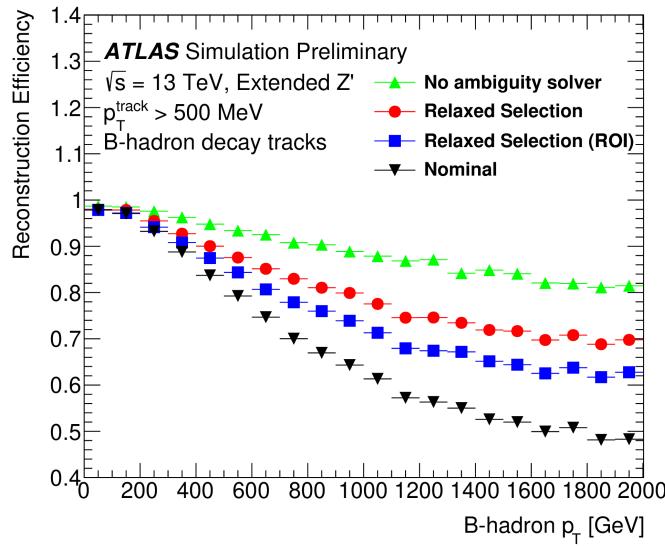


Figure 4.2: b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [103]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- p_T b -hadrons, 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.

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_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_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 at high p_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.

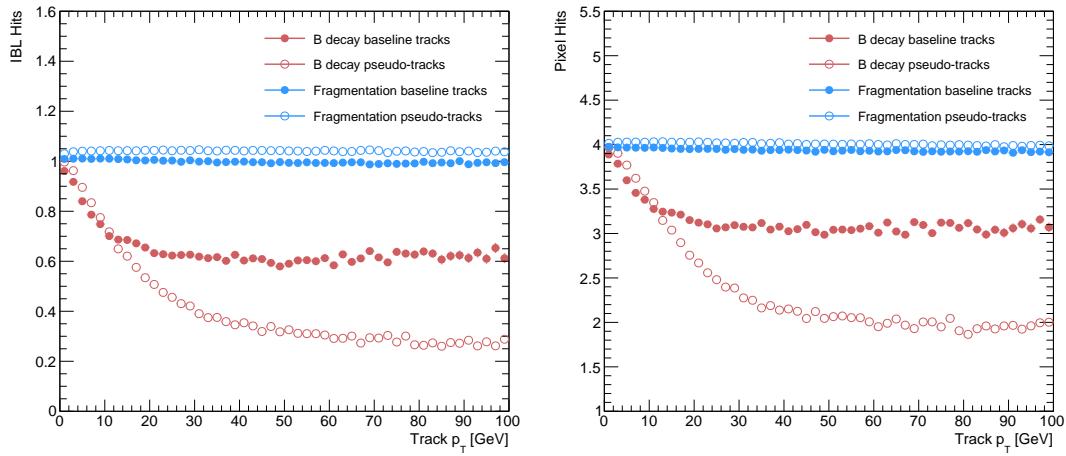


Figure 4.3: Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track for tracks in jets in a Z' sample at $\sqrt{s} = 13$ TeV. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in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

993 4.3 Investigations into High p_{T} b -hadron Tracking 994 Improvements

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

1000 4.3.1 Shared Hits

1001 The ambiguity solver is not run for pseudotracks. However, if the standard track
1002 collection is produced alongside the pseudotracks, then cluster splitting neural
1003 networks will be run for the standard tracks, and the resulting classification of
1004 clusters will be propagated to hits on pseudotracks. This quirk allows one to study
1005 the inefficiencies of the cluster splitting process, and relatedly to determine whether
1006 shared hit cuts in the ambiguity solver are too loose or too tight. The fraction of hits
1007 that are shared for the IBL and the B-layer is shown in Fig. 4.4. The shared hits on
1008 pseudotracks represent correctly assigned hits from merged clusters that were not
1009 able to be classified as split by the cluster splitting neural networks. As such, these
1010 represent the number of shared hits the ambiguity solver should aim to allow given
1011 the current performance of the cluster splitting algorithm. For shared hits on the IBL
1012 for particles produced before the IBL, the baseline selection appears to be successful
1013 in disallowing excessive numbers of shared hits. However, the ambiguity solver fails to
1014 limit shared hits for those particles produced after the IBL, reflecting the previously
1015 discussed problem of displaced tracks picking up incorrect hits. Meanwhile, it is clear
1016 that for the B-layer, the ambiguity solver is being overly aggressive in its rejection of
1017 shared hits.

1018 4.3.2 Global χ^2 Fitter Outlier Removal

1019 As part of the track fit an outlier removal procedure is run in which suspicious hits
1020 are identified and removed. This section documents ongoing studies into improving

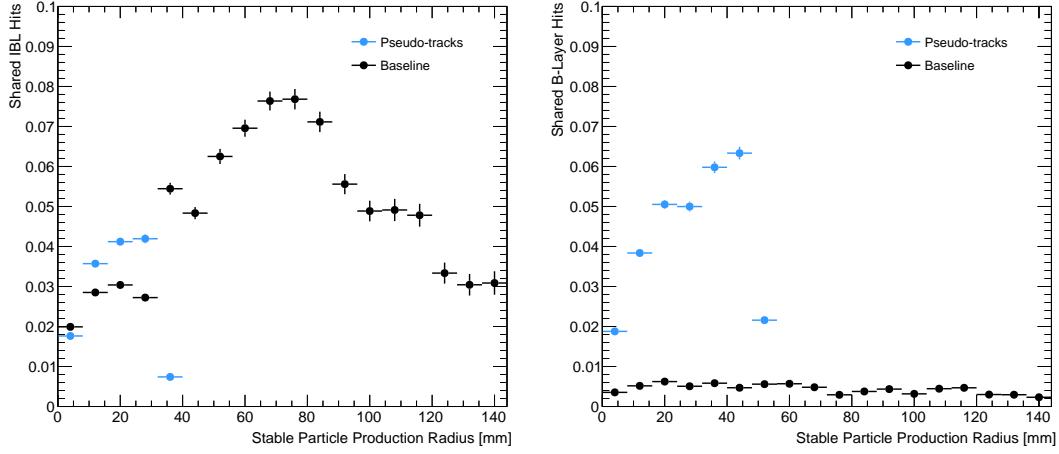


Figure 4.4: The fraction of IBL (left) and B-layer (right) hits which are shared on b -hadron decay tracks as a function of the production radius of the b -hadron decay product for tracks in jets in a Z' sample at $\sqrt{s} = 13$ TeV. Pseudotrack represent the ideal performance given the ATLAS detector,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

hit-to-track assignment by using the Global χ^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 [104] 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-

1038 detector which recorded the measurement (some sub-detectors are more precise than
1039 others). Additionally, a residual displacement $r_i = m_i - x_i$ between the predicted
1040 position of the track x_i (inclusive of the current measurement), and the position of
1041 the hit itself, m_i , is calculated. The pull p_i on the track state due to the current
1042 measurement is calculated according to

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

1043 This pull is computed for the transverse and longitudinal coordinates of the mea-
1044 surement, and the maximum of the two is selected and checked to see if it exceeds
1045 a certain selection threshold. If it does, the hit will be removed if the track also
1046 exceeds a threshold on the total χ^2/n , where n is the number of degrees of freedom
1047 on the track. The results of varying the outlier selection and χ^2/n thresholds are
1048 described below.

1049 Selection Optimisation

1050 A systematic variation of the outlier selection and χ^2/n thresholds has been carried
1051 out. Both thresholds were reduced in fixed step sizes of 0.25 for the outlier selection
1052 threshold and 1 for the χ^2/n threshold. The results for the best performing selections
1053 are discussed below. The value of the outlier selection threshold was reduced from 4
1054 down to 1.75, a change which affects the silicon layers (the TRT has separate outlier
1055 removal logic). Furthermore, a specific cut for the IBL was introduced, and after
1056 optimisation is set to 1.25. The second threshold on the track χ^2/n was also reduced
1057 from 7 to 4. Finally, instead of taking the maximum of the pulls in the longitudinal
1058 and transverse directions, a quadrature sum is taken of these two values and used.
1059 This variation is labelled “Mod GX2F” and was found to improve performance.

1060 The results are shown in Fig. 4.5 and demonstrate a reduction in wrong hit assignment
1061 whilst also improving slightly the rate at which good hits are assigned to tracks. For
1062 a 1 TeV track, the rate to assign good hits to the track increases by approximately
1063 10%, while the rate to assign incorrect hits decreases by approximately 16%. The
1064 improvements are also observed when looking inclusively in all tracks, which avoids
1065 the need for a specific b -jet region-of-interest selection.

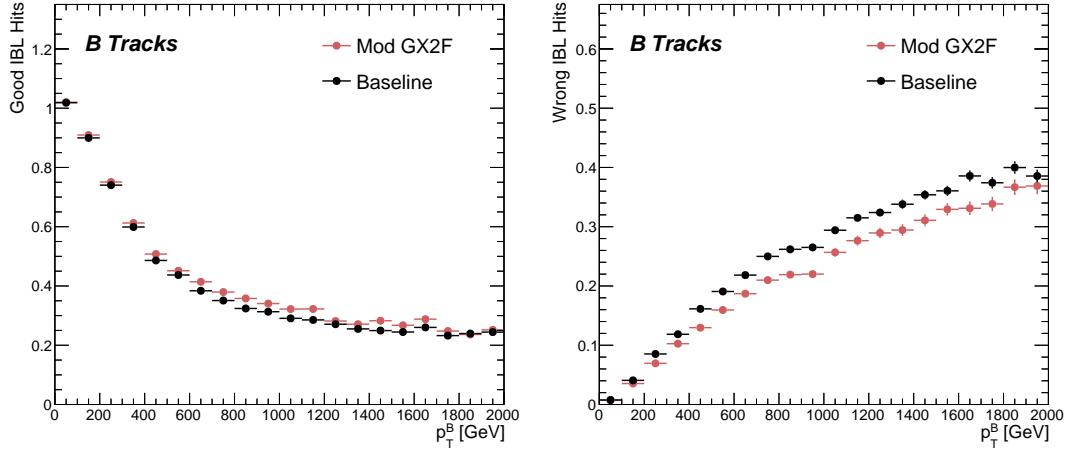


Figure 4.5: The average number of good (left) and wrong (right) IBL hits as a function of b -hadron p_T for tracks using baseline tracking (black) and the modified version of the outlier removal procedure (red).

1066 An improvement, though modest, of all track parameter resolutions and pulls is
 1067 observed. The improvement for the transverse impact parameter pull is shown in
 1068 Fig. 4.6. The results demonstrate an improvement in hit assignment, unchanged
 1069 reconstruction efficiency, and modest improvement in track parameter resolutions
 1070 and pulls. In addition, the truth match probability of tracks is unchanged, suggesting
 1071 that there is no increase in fake track rates. The changes are expected to have a
 1072 negligible impact on computational resources.

1073 4.4 Conclusion

1074 In this section, the difficulties facing efficient and accurate track reconstruction,
 1075 and hence performant b -tagging, have been outlined. The ambiguity solver, which
 1076 attempts to clean or reject tracks which have an excessive number of shared hits,
 1077 is shown to be overly aggressive in the removal of b -hadron decay product track
 1078 candidates. The ambiguity solving process relies on a complicated pre-defined
 1079 selection which has not been optimised for high transverse momentum b -hadron
 1080 track reconstruction. These conclusions have motivated further ongoing studies into
 1081 the improvement of the track reconstruction in dense environments and the high- p_T
 1082 regime, such as those in Ref. [103].

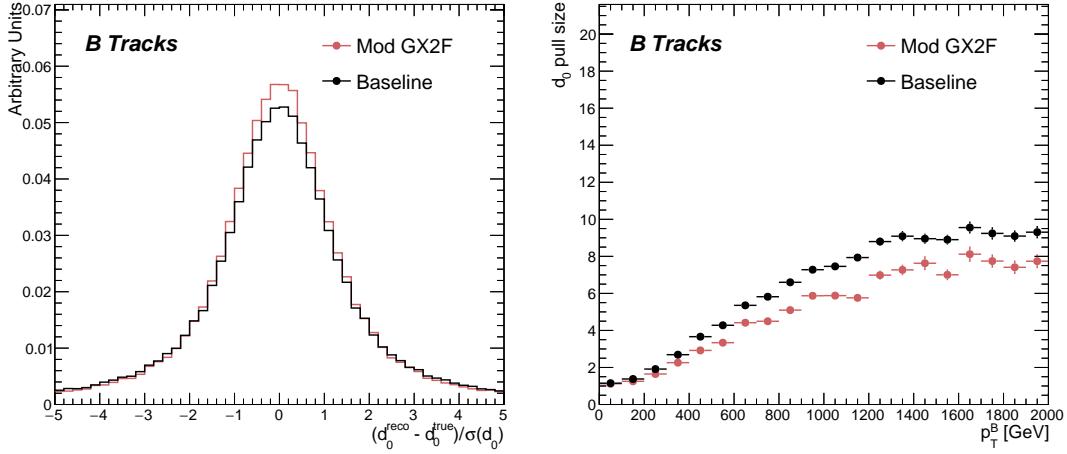


Figure 4.6: (left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified GX2F tracks. (right) The absolute value of the d_0 pull as a function of the truth b -hadron transverse momentum.

1083 An optimisation of the outlier removal process in the global χ^2 fitter was carried out.
 1084 The results of the optimisation show that more aggressive removal of outlier hits can
 1085 lead to fewer wrong hits being assigned to tracks, and improvements in the pulls of
 1086 the track parameters.

1087 Future Work

1088 The studies were carried out in Release 21 of the ATLAS software, and need to be
 1089 reproduced using the newer Release 22 to confirm the results against other changes
 1090 in the baseline tracking configuration. It is also necessary to study the impact of
 1091 the improved outlier removal on the downstream b -tagging algorithms. Thanks to
 1092 the all-in-one flavour tagging approach described in Chapter 6, this will in future be
 1093 easier to study.

1094 As there are some known data-MC discrepancies, fine tuned optimisation such as the
 1095 work presented here presents an opportunity to over-optimize the tracking algorithms
 1096 on MC. As such, further studies validating the improved outlier removal procedure
 1097 on data are required.

1098 **Chapter 5**

1099 **Track Classification MVA**

1100 This chapter details work on implementing a multivariate algorithm (MVA) to predict
1101 the truth origin of reconstructed tracks. An introduction to formalisms of machine
1102 learning is given in Section 5.1. In Section 5.2, the truth origin label is defined,
1103 and in Section 5.3 these labels are used to train a machine learning model that can
1104 effectively discriminate between good and fake tracks. Several studies motivated this
1105 work by demonstrating that at high p_T , b -tagging performance was degraded by the
1106 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm
1107 could be trained to detect fake tracks, these could be removed before their input to
1108 the b -tagging algorithms with the aim of improving performance. In addition, other
1109 groups that are sensitive to the presence of fake tracks would also benefit from this
1110 work.

1111 **5.1 Machine Learning Background**

1112 Over the past few decades, machine learning (ML) techniques have become increas-
1113 ingly popular in High Energy Physics experiments due the increased volumes of
1114 high-dimensional data and improvements in the techniques used (in particular deep
1115 learning). Machine learning is the process by which a computer program uses data
1116 to learn suitable parameters for a predictive model. This is opposed to explicitly
1117 providing instructions on how to perform a task. A subfield known as *supervised*
1118 *learning* is used in this work, and consists of exposing a model to a large number of
1119 labelled examples in order to extract relationships between the input data and their

1120 labels. These relationships are often complex, and explicitly programmed rules can
1121 fail to fully capture the relationships between inputs and outputs.

1122 In the simplest case, a set of m labelled training examples $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$
1123 is collected. Each element (x_i, y_i) consists of a input vector $x_i \in \mathbb{R}^{\text{input}}$, and the
1124 corresponding label y_i . In classification problems, these labels are integer *class labels*
1125 $y_i \in \{0, \dots, N - 1\}$, where N is the number of classes, which specify which of a
1126 pre-determined set of categorical classes the training example belongs to. The rest
1127 of the discussion in this chapter is limited to binary classification problems ($N = 2$).
1128 The two classes are often referred to as signal ($y_i = 1$) and background ($y_i = 0$),
1129 which need to be separated. Collecting sufficient and suitable data is one of the
1130 primary challenges of machine learning, as such data is not always readily available.
1131 Fortunately, sophisticated tools to simulate particle collisions have already been
1132 developed by the scientific community [105, 106]. These tools play a key role in
1133 generating a suitably large amount of labelled data which is used to train algorithms.
1134 More detail on the input datasets is given in Section 4.1.

1135 After obtaining suitable training data, the next step is to define a model. Given an
1136 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
1137 parameterised functional mapping from input space to output space. Given an input
1138 example x_i and a set of parameters θ , the model outputs a prediction $\hat{y}_i \in (0, 1)$ for
1139 the true label y_i , as in

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

1140 The output \hat{y}_i is in the interval $(0, 1)$ so as to be interpreted as the probability that
1141 the input example x_i belongs to the signal class. The parameters θ of the model are
1142 randomly initialised, and the model is designed to be expressive enough to correctly
1143 map the inputs x_i to the outputs y_i given a reasonable optimisation of the parameters.
1144 To perform this optimisation, the model is then trained, which amounts to showing
1145 the model a series of labelled training examples and modifying the parameters of the
1146 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 [107] and are easy to train via backpropagation [108].

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 [109, 110]. A diagram of an artificial neuron is shown in Fig. 5.1. Each neuron is defined by its parameters or *weights* θ 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 θ_0 . This term plays the role of a trainable constant value that is independent of the inputs.

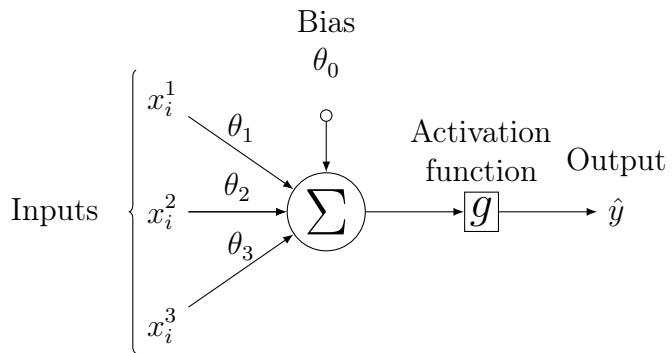


Figure 5.1: A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result z is passed to an activation function. Each neuron can be thought of as a logistic regression model.

The output of the dot product and bias term z 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) [111, 112] 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

1165 training. This is also why input data is typically normalised to have zero mean and
1166 unity variance [113].

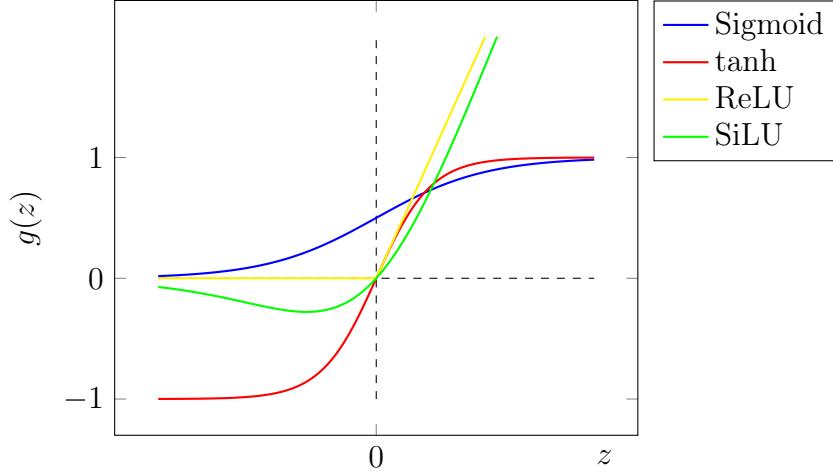


Figure 5.2: The output of several common choices for the activation function $g(z)$ of an artificial neuron. The input z is the output of the dot product between the activation and the weights, plus a bias term.

1167 Networks

1168 Several neurons are linked together in layers to form a neural network. The inputs
1169 are propagated layer-by-layer through the network until reaching the final output
1170 layer. The number of layers and neurons per layer are important hyperparameters
1171 (those parameters which are not optimised as part of the training process) which
1172 influence the performance of the model. In the case of binary classification, the final
1173 output layer generally consists of a single neuron with a sigmoid activation

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

1174 where z is the output from the dot product of the inputs and the weights, plus the
1175 bias term. This value is bounded between zero and one allowing the final output to
1176 be interpreted as the probability that the input sample belongs to the signal class.
1177 NNs have the crucial property of being differentiable functions, which facilitates the
1178 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 after 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 given in Eq. (5.3) loss is often used.

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

Since the model f is differentiable, a correction for each parameter θ_i can be computed by taking the partial derivative of L with respect to the parameter. Updated parameters θ'_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 α 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 [108]. 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 [114].

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 which a high 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^{match} , divided the total number of hits on the track N^{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 $\text{TMP} < 0.75$ ¹ 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

¹An 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 [57]. 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 Λ^0 decays, and hadronic interactions.

1230 shown for SV1 (see Section 3.4.2) in Fig. 5.4, where the efficiency to mistag a light-jet
 1231 decreases by up to 35% at a b -jet efficiency of 35% if such tracks are removed.

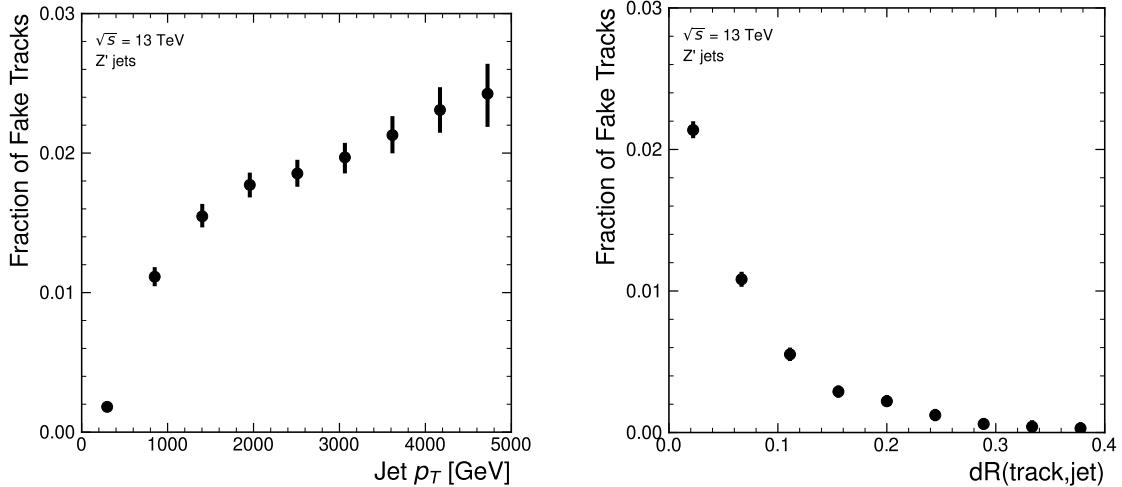


Figure 5.3: Rate of fake tracks as a function of jet transverse momentum (left) and $\Delta R(\text{track},\text{jet})$ (right) for jets in the Z' sample. The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases.

1232 To identify and remove fake tracks, a NN classification tool was trained with all
 1233 non-fake tracks as the signal class and fake tracks as the background class. Inputs to
 1234 the model are described in Section 5.3.1, while fake track removal performance is

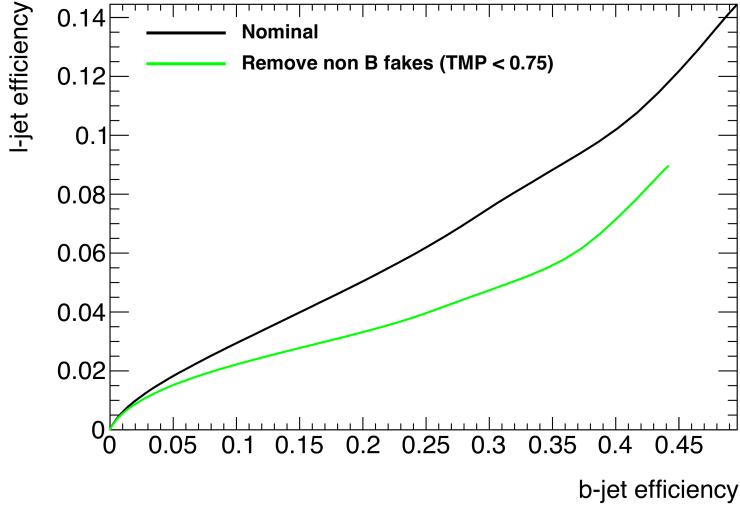


Figure 5.4: The light-jet efficiency of the low level tagger SV1 for jets in the Z' sample with $250 < p_T < 5000$ GeV, as a function of b -jet efficiency. The nominal tracking setup (black) is shown alongside the case where fake tracks which are not from the decay of a b -hadron are removed. The light-jet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

1235 given in Section 5.3.3. Both models are trained and evaluated using jets from the $t\bar{t}$
 1236 and Z' samples described in Section 4.1

1237 5.3.1 Model Inputs

1238 The fake track MVA is given two jet variables and 20 tracking related variables
 1239 for each track fed into the network. The jet transverse momentum and signed
 1240 pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in
 1241 Table 5.2.

1242 The track parameters and hit pattern are key indicators of whether or not a track
 1243 is fake. The FracRank variable is the ordered index of the tracks that pass the
 1244 ambiguity solver's selection divided by the total number of successfully reconstructed
 1245 tracks in the event. The ambiguity solver processes track candidates iteratively
 1246 in order of an internal score (see Section 3.4.1), and the order in which tracks are
 1247 accepted is preserved. Since tracks with shared hits have lower scores, tracks which
 1248 do not require the removal of shared hits are likely to be processed and accepted

Jet Input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	Description
p_T	Track transverse momentum
ΔR	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_{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 [57], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.

1249 earlier on, whereas tracks with shared hits will be processed later and potentially
1250 have their shared hits removed. Hence the FracRank variable gives an indication
1251 of the track quality and how likely it is that hits would have been removed (tracks
1252 processed later on are more likely to have hits removed).

1253 Track selection follows the loose selection described in Ref. [87] and outlined in
1254 Table 3.2, which was found to improve the performance compared to previous tighter
1255 selections, whilst ensuring good resolution of the track’s parameters and a low fake
1256 rate [57]. Inputs are scaled to have a central value of zero and a variance of unity
1257 before training and evaluation.

1258 5.3.2 Model Hyperparameters

1259 Due to the imbalance between the two classes (with fake tracks being relatively
1260 uncommon), a weight was added to the loss function for the background class to
1261 balance their relative weights. The NN was made up of two hidden layers with
1262 220 nodes per layer. The ReLU activation function was used in conjunction with
1263 the Adam optimiser with a learning rate of $1e-3$. Optimisation of the networks
1264 architecture was carried out to ensure optimal performance with a relatively small
1265 number of learnable parameters – 54,000. The model was trained using 40 million
1266 tracks with a further 4 million tracks each used for validation and testing. A full list
1267 of the model hyperparameters is given in Table 5.3.

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: Hyperparameter for the track classification model

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 MVA. 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 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.

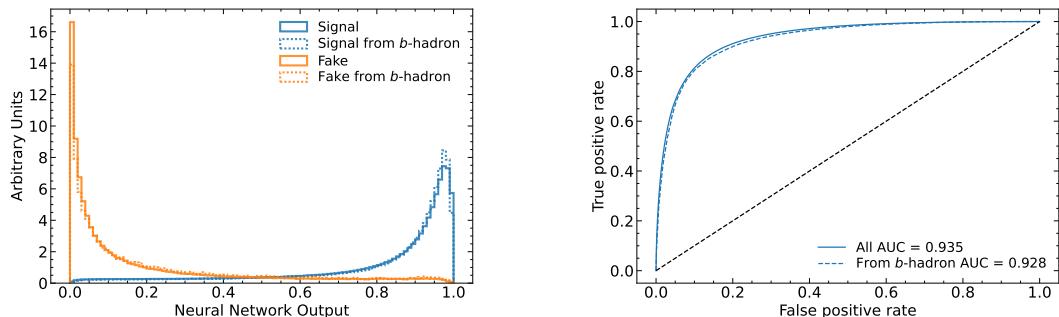


Figure 5.5: (left) Normalised histograms of the fake track classification model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed 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

1286 happen because fake tracks with $\text{TMP} < 0.75$ are still matched to a truth particle,
1287 which can be the decay product of a b -hadron.

MVA Output Cut	Signal Track Efficiency		Fake Track Efficiency	
	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.

1288 5.4 b -hadron Track Identification

1289 After initial tests and investigation, it was found that fake tracks which were the
1290 result of b -hadron decays actually aided b -tagging performance, as demonstrated in
1291 Fig. 5.7. The application of a single tool which removed all fake tracks was therefore
1292 not optimal. A second tool was therefore trained in the same manner as the first,
1293 this one was designed to distinguish between those tracks which were from the decay
1294 of a b -hadron (FromB and FromBC in Table 5.1) and those which were not (all other
1295 truth origins). Fake tracks which were from the decay of a b -hadron were included in
1296 the signal class. The b -hadron decay track MVA was trained using the same setup
1297 as described above, with the same tracks, input variables, and training procedure.
1298 The performance of the model to separate b -hadron decay tracks from other tracks
1299 is shown in Fig. 5.6. Using a selection WP of 0.1, the model can retain 98.5% of
1300 b -hadron tracks and reject 46.2% of tracks not from the decay of a b -hadron. In
1301 Section 5.5, this model is used in conjunction with the fake track identification MVA
1302 to identify and remove fake tracks which are not from the decay of a b -hadron.

1303 5.5 Combined Approach

1304 A 2-dimensional cut was then used to only reject those tracks that had a high
1305 probability of being fake, and also a low probability of being a b -hadron decay track.
1306 The results of the combined approach are provided in Table 5.5, which shows that

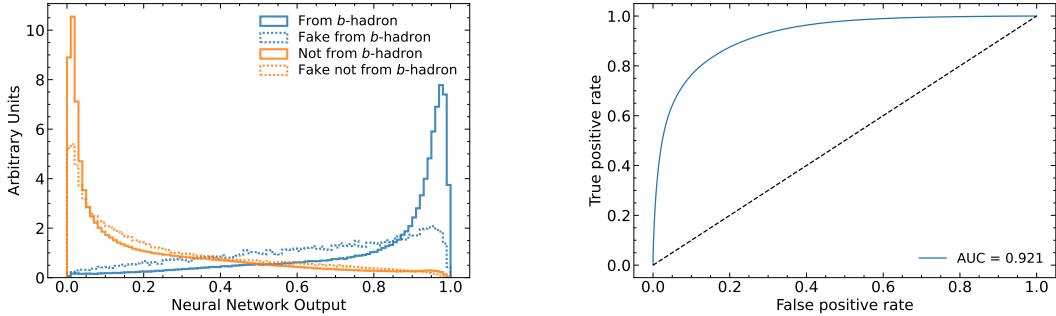


Figure 5.6: (left) Normalised histogram of the b -hadron track identification model output separated for tracks from the decay of a b -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line). The plots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

1307 for the working point “A”, 98.6% of b -hadron decay tracks (both good and fake) are
 1308 retained, while 50.7% of fake tracks which are not from b -hadron decays are rejected.

WP	Fake MVA Cut	b -hadron Decay MVA 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 MVAs for the two defined working points. Working point “B” cuts more aggressively on the MVA 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).

1309 The light-jet efficiency of SV1 is successfully reduced when using the combined tools
 1310 to remove fake tracks that are not from a b -hadron decay, as shown in Fig. 5.7. At a
 1311 b -jet efficiency of 70%, the light-jet mistag rate for jets with $250 < p_T < 400$ GeV
 1312 is reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For
 1313 jets with $400 < p_T < 1000$ GeV the mistag rate drops from 0.1 to 0.08 for a similar
 1314 relative improvement of 20%. The performance of the fake track removal approach
 1315 was also tested for the other low level vertexing algorithm – JetFitter. A similar level
 1316 of improvement in the light-jet mistag rate was observed with a reduction of up to a
 1317 20% reduction for both low- and high- p_T jets in the Z' sample achieved. Together,
 1318 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.

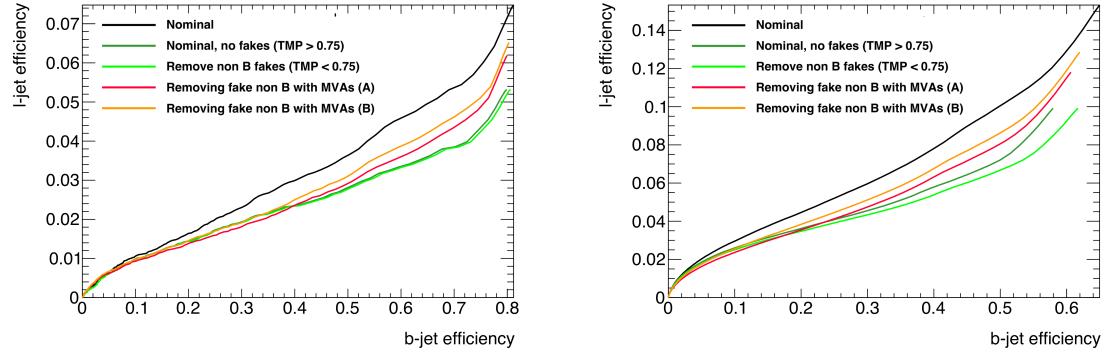


Figure 5.7: The effect of applying the fake track identification algorithm together with the b -hadron decay track identification on the jet tagging performance of SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (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 b -tagging performance. A ML tool to identify fake tracks has been developed, which can be used to limit the number of fake tracks being input to the b -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 in general comparable with that achieved when using the truth information to remove the fake tracks not from the decay of a b -hadron.

1336 Future Work

1337 While removing tracks prior to their input to the low level tagging algorithms is
1338 shown here to be beneficial, a more performant alternative might be to keep these
1339 tracks but label them as being fake (for example using the output of the classification
1340 tool), and allow the tagging algorithms to take this into consideration. This is not
1341 straightforward with manually optimised taggers such as SV1 and JetFitter, but is
1342 possible with more advanced taggers as described in Chapter 6.

1343 Tools which identify the origin of a given track have other potential uses. One
1344 application is to isolate a relatively pure sample of fake tracks which can be used
1345 to estimate the fake track rate in data, which would be useful for estimating the
1346 uncertainty on fake track modelling. Another application is to use the b -hadron
1347 track identification tool to improve the track-to-jet association. Both applications
1348 are currently under investigation.

1349 The approach here works on a track-by-track basis, but a more sophisticated approach
1350 would consider the correlations between the tracks inside a jet. Also left for future
1351 work is to simultaneously train a single tool which discriminates between all the
1352 truth origins listed in Table 5.1. Such a tool would be useful as a general purpose
1353 multiclass classifier. An algorithm which takes both these aspects into consideration
1354 is discussed in Chapter 6.

₁₃₅₅ **Chapter 6**

₁₃₅₆ **Graph Neural Network Flavour
Tagger**

₁₃₅₈ This chapter introduces GN1, a novel ML-based flavour tagging algorithm based on
₁₃₅₉ graph neural networks (GNNs). In Section 6.1, an overview of the approach used
₁₃₆₀ for GN1 is given. An introduction to the theory of GNNs is provided in Section 6.2.
₁₃₆₁ Details of the experimental setup are provided in Section 6.3, while the architecture
₁₃₆₂ of GN1 is specified in Section 6.4.3. In Section 6.4.4, the training procedure is
₁₃₆₃ described, and in Section 6.5 the results are shown.

₁₃₆₄ **6.1 Motivation**

₁₃₆₅ GN1 is a monolithic approach to flavour tagging as illustrated in Fig. 6.1. As opposed
₁₃₆₆ to the existing approach to flavour tagging described in Chapter 4, which relies
₁₃₆₇ on a two tiered approach requiring the use of both low- and high-level algorithms,
₁₃₆₈ GN1 takes as inputs information directly from an unordered variable number of
₁₃₆₉ tracks as input, and predicts the jet flavour without requiring outputs from the
₁₃₇₀ intermediate low-level algorithms. In addition to predicting the flavour of the jet, the
₁₃₇₁ model predicts which physical processes produced the various tracks, and groups the
₁₃₇₂ tracks into vertices. These auxiliary training objectives provide valuable additional
₁₃₇₃ information about the contents of the jet and enhance the performance of the primary
₁₃₇₄ flavour prediction task. The use of GNNs offers a natural way to classify jets with

1375 variable numbers of unordered associated tracks (see Section 6.2), while allowing for
1376 the inclusion of auxiliary training objectives [2, 115].

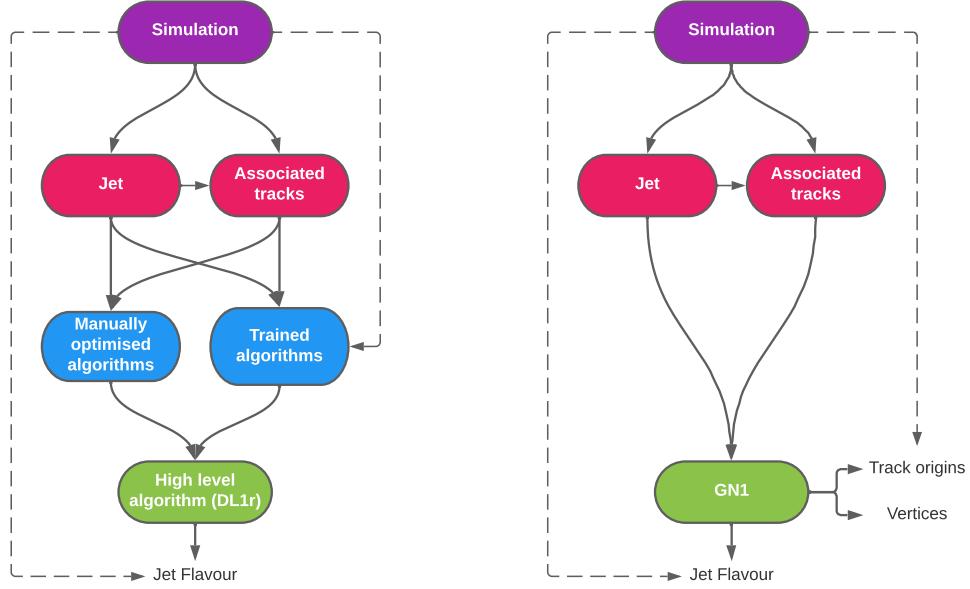


Figure 6.1: Comparison of the existing flavour tagging scheme (left) and GN1 (right). The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, additional 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].

1377 The current flavour tagging algorithms utilise a two-tired approach, with five low
1378 level algorithms feeding intermediate features into the high-level tagger DL1r, which
1379 outputs variables which discriminate between the different jet flavours. In contrast
1380 GN1 consists of only a single neural network, which takes the tracks as inputs along
1381 with some kinematic information about the jet. As a result, it does not depend on
1382 the outputs of any other flavour tagging algorithm. A simple training of the model
1383 fully optimises its parameters, representing a significant simplification with respect to
1384 the optimisation procedure for DL1r. This is particularly important when optimising
1385 the tagger for new regions of phase space (e.g. c -tagging or high- p_T b -tagging), or
1386 when the detector is upgraded or the charged particle reconstruction or selection
1387 algorithms are re-optimised.

1388 GN1 is trained to learn about the internal structure of the jet through the use of two
1389 auxiliary training objectives: the prediction of the underlying physics process from
1390 which each track originated, and the grouping of tracks originating from a common
1391 spatial position (i.e. a common vertex). These auxiliary objectives are meant to
1392 guide the neural network towards a more complete understanding of the underlying
1393 physics inside the jet, thereby removing the need for the low-level algorithms, which
1394 previously contained information about the underlying physics in their design. The
1395 training targets for the primary and auxiliary objectives are extracted from truth
1396 information, as opposed to reconstructed quantities available in both collision data
1397 and simulation.

1398 In this chapter, the following advantages of the GN1 approach will be demonstrated:

- 1399 1. GN1 boasts improved performance with respect to the current ATLAS flavour
1400 tagging algorithms, with significantly larger background rejection rates for a
1401 given signal efficiency. Alternatively the rejection rates can be kept fixed for a
1402 substantial increase in signal efficiency, in particular at high- p_T .
- 1403 2. The same network architecture can be easily optimised for a wider variety of
1404 use cases (e.g. c -jet tagging and high- p_T jet tagging) since there are no low-level
1405 algorithms to retune.
- 1406 3. There are fewer algorithms to maintain.
- 1407 4. Alongside the network's prediction of the jet flavour, the auxiliary vertex and
1408 track origin predictions provide more information on why a jet was (mis)tagged
1409 or not. This information can also have uses in other applications, for instance
1410 to explicitly reconstruct displaced decay vertices or to remove fake tracks.¹

1411 6.2 Graph Neural Network Theory

1412 Graph neural networks are a more sophisticated neural network model (see Sec-
1413 tion 5.1.1) that are designed to operate on graph structured data. A brief introduction
1414 to GNNs is provided in this section following the formalism in Ref. [116].

¹A 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. [57].

1415 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 $\mathcal{E} =$
1416 $\{e_i\}_{i=1:N^e}$, and a global representation u . Each node represents an individual object,
1417 and edges are directed connections between two nodes, called the *sender* and *receiver*
1418 nodes. The connectivity of the graph therefore encodes information about the
1419 relationships between objects that exist in the graph.

1420 A single graph network layer consists of three separate update functions ϕ^e , ϕ^h and
1421 ϕ^u one for each of the nodes, edges, and global graph representation, and similarly
1422 three aggregation functions $\rho^{e \rightarrow h}$, $\rho^{e \rightarrow u}$ and $\rho^{h \rightarrow u}$. The aggregation functions combine
1423 information across different edges or nodes for input into the update functions,
1424 which produce new representations for the nodes, edges and global objects based on
1425 the information in the previous layer and the aggregated information. The update
1426 functions are typically each implemented as a dense feedforward neural network (as
1427 described in Section 5.1.1). The edges e_i are updated by a edge network ϕ^e as in

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

1428 where h_s and h_r are the sender and receiver nodes respectively. The nodes are
1429 updated with a node network ϕ^h as in

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

1430 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
1431 referred to as the edge aggregation function. The global representation is updated
1432 using the global network ϕ^u as in

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

1433 where \bar{e}' is the aggregation $\rho^{e \rightarrow u}$ over all updated edges e'_i and \bar{h}' is the aggregation
1434 $\rho^{h \rightarrow u}$ over all updated nodes h'_i .

1435 The graph network layer performs a graph convolution, in an analogous way to a
1436 convolutional neural network operating on a grid of pixels. The above description
1437 is general, and not all concrete implementations of GNNs need to implement every
1438 aspect. For example, the global graph representation need not be present, and
1439 it is also possible that no dedicated edge features are present. In such cases the
1440 corresponding update and aggregation functions are not needed. Fig. 6.2 shows two

1441 possible graph network update layers. The layer used in the GN1 model is specified
1442 in more detail in Section 6.4.3.

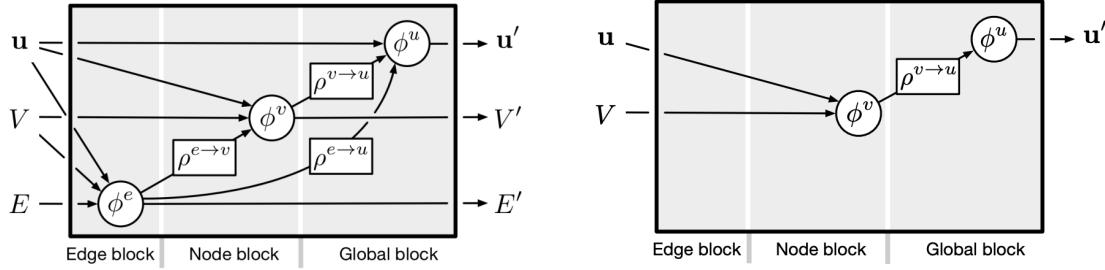


Figure 6.2: The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [117]. Reproduced from Ref. [116].

1443 6.3 Experimental Setup

1444 6.3.1 Datasets

1445 The datasets used to train the GN1 tagger are the same as described in Section 4.1.
1446 The training dataset contains 30 million jets, 60% of which are $t\bar{t}$ jets and 40% of
1447 which are Z' jets. In order to evaluate the performance of the model during training,
1448 a statistically independent validation set of 500k jets from both the $t\bar{t}$ and Z' samples
1449 are used. For the testing of the model and the creation of the performance plots,
1450 a further 1 million independent testing jets from each of the $t\bar{t}$ and Z' samples are
1451 used. Before being fed into the model, the track- and jet-level inputs are normalised
1452 to have a mean of zero and a variance of unity. The jet flavour labels are assigned
1453 as described in Section 3.4.3. Truth labelled b -, c - and light-jets are kinematically
1454 re-sampled in p_T and η to ensure identical distributions in these variables.

1455 6.4 Model Architecture

1456 6.4.1 Model Inputs

1457 As inputs, GN1 is fed two kinematic jet variables and an unordered set of up to 40
1458 tracks which have been associated to the jet. If more than 40 tracks are associated
1459 to a given jet, only the first 40 tracks with the largest transverse IP significance $s(d_0)$
1460 (see Section 3.2.2) are fed into the model as inputs. Each track is characterised by 21
1461 variables as detailed in Table 6.1. The kinematic jet variables are the jet transverse
1462 momentum and signed pseudorapidity. For each track, variables containing the
1463 track parameters and uncertainties, and detailed information on the hit content are
1464 provided as inputs to the model.

1465 Dependence of the model on the absolute value of the azimuthal jet angle ϕ is
1466 explicitly removed by providing only the azimuthal angle of tracks relative to the jet
1467 axis. The track pseudorapidity is also provided relative to the jet axis.

1468 Since heavy flavour hadrons can decay semileptonically approximately 40% of the time,
1469 the presence of a reconstructed lepton in the jet carries discriminating information
1470 about the jet flavour. To exploit this, a variant of GN1 called GN1Lep is trained in
1471 addition to the baseline model. The GN1Lep variant is identical to the baseline model,
1472 except for the inclusion an additional track-level input, leptonID, which indicates
1473 if the track was used in the reconstruction of an electron, a muon or neither. The
1474 variable is signed by the charge of the reconstructed lepton. The leptons used in the
1475 definition of the leptonID variable are required to satisfy basic quality requirements.
1476 The muons are required to be combined [118], and the electrons are required to pass
1477 the *VeryLoose* likelihood-based identification working point [119].

1478 The selections applied to the tracks, outlined in Table 3.2, is the same as that used
1479 for the fake track classification MVA described in Chapter 5. However, Section 6.6.1
1480 demonstrates that further relaxation of the track selection requirements may be
1481 warranted.

Jet Input	Description
p_T	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	Description
q/p	Track charge divided by momentum
$d\eta$	Pseudorapidity of the track, relative to the jet η
$d\phi$	Azimuthal angle of the track, relative to the jet ϕ
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 θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	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 [57], 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].

1482 6.4.2 Auxiliary Training Objectives

1483 In addition to the jet flavour classification, two auxiliary training objectives are
1484 defined. The first auxiliary objective is the prediction of the physical process that
1485 gave rise to each track within the jet (i.e. the track origin), while the second is the
1486 prediction of track-pair vertex compatibility. Each auxiliary training objective comes
1487 with a training target which, similar to the jet flavour label, is a truth labels derived
1488 from the simulation. The presence of the auxiliary training objectives improves the
1489 jet classification performance as demonstrated in Section 6.5.3.

1490 For the track origin prediction objective, each track is labelled with one of the
1491 exclusive categories defined in Table 5.1 of Section 5.2 after analysing the particle
1492 interaction (or lack thereof) which led to its formation. Since the presence of different
1493 track origins is strongly related to the flavour of the jet, training GN1 to recognise
1494 the origin of the tracks provides an additional handle on the classification of the
1495 jet flavour. This task may also aid the jet flavour prediction by acting as a form of
1496 supervised attention [120] - in detecting tracks from heavy flavour decays the model
1497 may learn to pay more attention to these tracks.

1498 The vertexing auxiliary objective makes use of the fact that displaced decays of b -
1499 and c -hadrons lead to secondary and tertiary vertices inside the jet, as described
1500 in Section 3.5.2. The presence of displaced secondary vertices is not a completely
1501 clean signal of a heavy flavour jet, as displaced secondary vertices can also occur
1502 in light-jets as a result of material interactions, conversions, and long-lived particle
1503 decays (e.g. K_S^0 and Λ^0). For this objective, GN1 predicts a binary label for each pair
1504 of tracks in the jet. The label has a value of 1 if the truth particles associated with
1505 the two tracks in the pair originated from the same spatial point, and 0 otherwise. To
1506 derive the corresponding truth labels for training, truth production vertices within
1507 0.1 mm are merged. Track-pairs where one or both of the tracks in the pair have an
1508 origin label of either Pile-up or Fake are assigned a label of 0. Using the pairwise
1509 predictions from the model, groups of tracks that have common compatibility can
1510 be formed, resulting in the identification of vertices. Two existing low-level tagging
1511 algorithms, SV1 and JetFitter (introduced in Section 3.4.2), are currently used to
1512 find and reconstruct vertices inside jets and are used as inputs to the existing jet
1513 flavour tagger DL1r. The addition of this auxiliary training objective removes the
1514 need for inputs from a dedicated secondary vertexing algorithm.

1515 Both of the auxiliary training objectives described here can be considered as “stepping
1516 stones” on the way to classifying the flavour of the jet. By requiring the model to
1517 predict the truth origin of each track and the vertex compatibility of each track-pair,
1518 the model is guided to learn representations of the jet which are connected to the
1519 underlying physics and therefore relevant for classifying the jet flavour.

1520 **6.4.3 Architecture**

1521 A coarse optimisation of the network architecture hyperparameters (for example
1522 number of layers and number of neurons per layer) has been carried out in order
1523 to maximise the flavour tagging performance, but it is likely that further dedicated
1524 optimisation studies could lead to further performance improvements.

1525 The model architecture builds on a previous implementation of a GNN-based jet
1526 tagger [115]. The previous approach was comprised of two separate graph neural
1527 networks with the auxiliary tasks being performed at an intermediate stage after the
1528 first and before the second. This two stage approach was found to be unnecessary and
1529 as such GN1 simplifies the architecture into a single graph neural network with the
1530 auxiliary tasks being performed at the end, alongside the primary jet classification
1531 task. GN1 makes use of a more sophisticated graph neural network layer [121],
1532 which is described in more detail below. The changes significantly improved tagging
1533 performance and also led to a significant reduction in training time.

1534 As inputs, the model takes information about the jet and a number of associated
1535 tracks, as detailed in Section 6.4.1. The jet variables are concatenated with the
1536 variables for each track as shown in Fig. 6.3. The combined jet-track input vectors
1537 are then fed into a per-track initialisation network with three hidden layers, each
1538 containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 6.4. The
1539 track initialisation network is similar to a deep sets model [117], but does not include
1540 a reduction operation (mean or summation) over the output track representations.
1541 The initialisation network allows for initial per-track input processing without the
1542 associated parameter count cost of the graph convolutional layers described below.

1543 The outputs of the track initialisation network are used to populate the nodes of a
1544 fully connected graph, such that each node in the graph neighbours every other node.
1545 Each node h_i in the graph corresponds to a single track in the jet, and is characterised

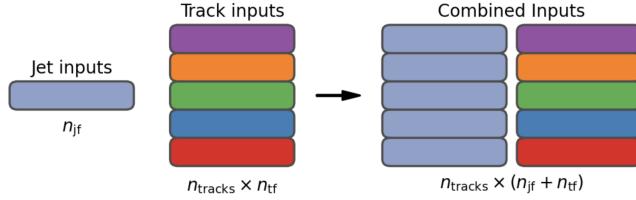


Figure 6.3: The inputs to GN1 are the two jet features ($n_{jf} = 2$), and an array of n_{tracks} , where each track is described by 21 track features ($n_{tf} = 21$). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1 [3].

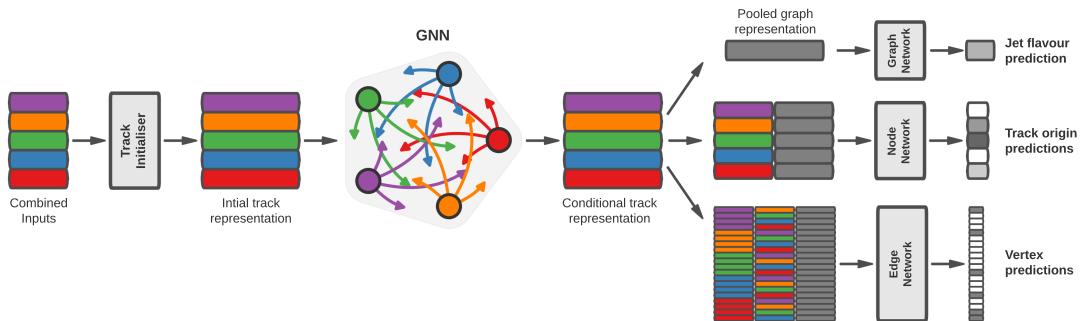


Figure 6.4: The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting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

1546 by a feature vector, also called a representation. The per-track output representations
 1547 from the initialisation networks are used as the initial feature vectors of each node
 1548 in the graph. In each layer of the graph network, output node representations h'_i
 1549 are computed by aggregating the features of h_i and neighbouring nodes \mathcal{N}_i using
 1550 a multi-head attention mechanism ($n = 2$) as described in Ref. [121, 122]. First,
 1551 the feature vectors of receiver and sender nodes are fed into two fully connected
 1552 linear layers \mathbf{W}_r and \mathbf{W}_s , to produce an updated representation for each sender and
 1553 receiver node $\mathbf{W}_r h_i$ and $\mathbf{W}_s h_j$. These updated feature vectors are used to compute
 1554 edge scores $e(h_i, h_j)$ for each node pair, as in

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

1555 where, θ is a non-linear activation function, and \mathbf{a} is a learned vector. These edge
 1556 scores are then used to calculate attention weights a_{ij} for each pair of nodes using
 1557 the softmax function over the edge scores

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

1558 Finally, the updated representations for the receiver nodes h'_i are computed by taking
 1559 the weighted sum over each updated node representation $\mathbf{W}_r h_i$, with weights a_{ij}

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

1560 The set of operations described above constitute a single graph network layer. Three
 1561 such layers are stacked to construct the graph network, representing a balance
 1562 between achieving good performance in a reasonable time and avoiding overtraining
 1563 due to inflation of the parameter count of the model. The final output from the graph
 1564 neural network is a set of per-node (i.e. per-track) feature vectors that are conditional
 1565 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 [111].

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

1590 network is comprised of four fully connected hidden layers with 128, 64, 32 and 16
1591 neurons respectively, and has three outputs corresponding to the b -, c - and light-
1592 jet classes. To obtain probability outputs for each task, the outputs from each
1593 classification network are passed through a softmax function.

1594 6.4.4 Training

1595 The full GN1 training procedure minimises the total loss function L_{total} , defined as

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

1596 This loss is composed of three terms: L_{jet} , the categorical cross entropy loss over the
1597 different jet flavours; L_{vertex} , the binary track-pair compatibility cross entropy loss;
1598 and L_{track} , the categorical cross entropy loss for the track origin prediction. L_{vertex} is
1599 computed via a weighted average over all intra-jet track-pairs in the batch, and L_{track}
1600 is computed by a weighted average over all tracks in the batch, where the weights
1601 are described below.

1602 The different losses converge to different values during training, reflecting differences
1603 in the relative difficulty of the various training objectives. The values of L_{vertex} and
1604 L_{track} are weighted by $\alpha = 1.5$ and $\beta = 0.5$ respectively to ensure they converge to
1605 similar values, giving them an equal weighting towards L_{total} . The values of α and
1606 β are chosen to ensure that L_{jet} converges to a larger value than either L_{vertex} and
1607 L_{track} , which reflects the primary importance of the jet classification objective. It
1608 was found that in practice the overall performance of the model was not sensitive
1609 to modest changes in the loss weights α and β . Pre-training using L_{total} (i.e. on
1610 all tasks) and fine tuning on only the jet classification task also did not improve
1611 performance versus the standard setup, indicating that the auxiliary tasks are not
1612 in direct competition with the jet classification task. As there was a large variation
1613 in the relative abundance of tracks of the different origins, the contribution of each
1614 origin to L_{track} was weighted by the inverse of the frequency of their occurrence. In
1615 vertexing loss L_{vertex} , the class weight for track-pairs where both tracks are from
1616 either a b - or c -hadron was increased by a factor of two as compared with other

1617 track-pairs, to encourage the network to focus on correctly classifying heavy flavour
1618 vertices.

1619 GN1 can be trained with either the node or edge networks (and their corresponding
1620 auxiliary tasks), or both, removed, as discussed in Section 6.5.3. In such cases,
1621 the corresponding losses L_{vertex} and L_{track} are also removed from the calculation
1622 of the overall loss L_{total} . The performance of the resulting models provides a
1623 useful indication of the benefit of including the auxiliary tasks to the primary jet
1624 classification objective.

1625 GN1 was trained for 100 epochs on 4 NVIDIA V100 GPUs. Each epoch takes
1626 approximately 25 mins to complete over the training sample of 30 million jets. The
1627 Adam optimiser [123] with an initial learning rate of $1e-3$, and a batch size of 4000
1628 jets (spread across the 4 GPUs) was used. Typically the validation loss, calculated
1629 on 500k jets, became stable after around 60 epochs. The epoch that minimized the
1630 validation loss was used for evaluation. GN1 has been integrated into the ATLAS
1631 software [54] using ONNX [124]. The test sample jet flavour predictions scores are
1632 computed using the ATLAS software stack as a verification of this process.

1633 6.5 Results

1634 The GN1 tagger is evaluated both as a b -tagging and c -tagging algorithm in Sec-
1635 tion 6.5.1 and Section 6.5.2 respectively. Evaluation is performed separately on
1636 jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ and jets in the Z' sample with
1637 $250 < p_T < 5000 \text{ GeV}$. The performance of the model is compared to the DL1r
1638 tagger [89, 90], which has been retrained on 75 million jets from the same samples as
1639 GN1. The input RNNIP tagger [86] to DL1r has not been retrained. As discussed,
1640 each tagger predicts the probability that a jet belongs to the b -, c - and light-classes.
1641 To use the model for b -tagging, these probabilities are combined into a single score
1642 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. [89]. 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.

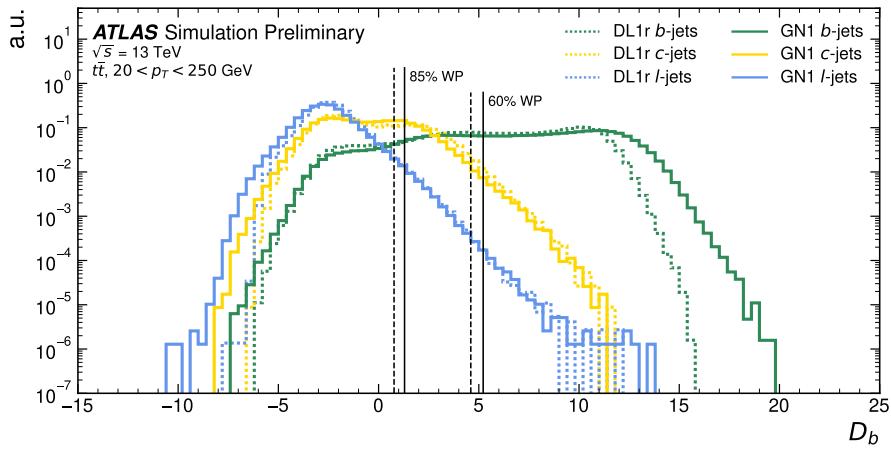


Figure 6.5: Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of $f_c = 0.018$ is 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].

6.5.1 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

1659 typically used by analyses, the corresponding c - and light-jet rejection rates are
1660 displayed in Figs. 6.6 and 6.7 for jets in the $t\bar{t}$ and Z' samples respectively. Four
1661 standard WPs are defined with b -jet tagging efficiencies of 60%, 70%, 77% and 85%
1662 respectively. These WPs are commonly used by physics analyses depending on their
1663 specific signal and background requirements. The WPs are defined based on jets in
1664 the $t\bar{t}$ sample only. Due to the much higher jet p_T range in the Z' sample, and the
1665 increased difficulty in tagging jets at high- p_T (see Chapter 4), the corresponding b -jet
1666 tagging efficiencies for jets in the Z' sample are lower than the corresponding WPs
1667 calculated in the $t\bar{t}$ sample. For instance the WP cut value computed to provide a
1668 70% b -jet tagging efficiency on the $t\bar{t}$ sample results in a b -jet tagging efficiency of
1669 just $\sim 30\%$ on the Z' sample. In order to account for this, the range of b -jet tagging
1670 efficiencies displayed for plots showing the performance for jets in the Z' sample (for
1671 example Fig. 6.7) is chosen to span the lower efficiencies achieved in the Z' sample
1672 at high- p_T .

1673 For jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$, GN1 demonstrates considerably
1674 better c - and light-jet rejection when compared with DL1r across the full range of
1675 b -jet tagging efficiencies studied. The relative improvement is strongly dependent
1676 on the b -jet tagging efficiency under study. The largest improvements are found at
1677 lower b -jet tagging efficiencies. At a b -jet tagging efficiency of 70%, the c -jet rejection
1678 improves by a factor of ~ 2.1 while the light-jet rejection improves by a factor of ~ 1.8
1679 with respect to DL1r. For high- p_T jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$,
1680 GN1 also brings a significant performance improvement with respect to DL1r across
1681 the range of b -jet tagging efficiencies studied. Again, the largest relative improvement
1682 in performance comes at the lower b -jet tagging efficiencies. At a b -jet efficiency of
1683 30%, GN1 improves the c -jet rejection with respect to DL1r by a factor of ~ 2.8 and
1684 the light-jet rejection by a factor of ~ 6 . The performance comparison at lower b -jet
1685 tagging efficiencies is made more difficult due to the increased statistical uncertainties
1686 which result from the high rejection of background.

1687 The GN1Lep variant of GN1 demonstrates further improved performance with respect
1688 to the baseline model. This demonstrates the additional jet flavour discrimination
1689 power provided by the leptonID track input. For jets in the $t\bar{t}$ sample, the relative c -
1690 jet rejection improvement with respect to GN1 at the 70% b -jet WP is approximately
1691 25%. The improvement in light-jet rejection also increases by 40% at the same WP.
1692 For jets in the Z' sample, the relative c -jet rejection (light-jet rejection) performance

1693 with respect to GN1 improves by approximately 10% (25%) at a b -jet tagging
1694 efficiency of 30%.

1695 In general, the performance of all the taggers is strongly dependent on jet p_T , due
1696 to the increased multiplicity and collimation of tracks, and the displaced decays
1697 that result from within the heavy flavour jets (see Chapter 4). Together, they
1698 contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a
1699 general degradation in quality of tracks inside the core of a jet, which in turn reduces
1700 the jet tagging performance. In order to study how the tagging performance changes
1701 as a function of the jet p_T , the b -jet tagging efficiency as a function of p_T for a fixed
1702 light-jet rejection of 100 in each bin is shown in Fig. 6.8. For jets in the $t\bar{t}$ sample,
1703 at a fixed light-jet rejection of 100, GN1 improves the b -jet tagging efficiency by
1704 approximately 4% across all the jet p_T bins. Meanwhile, GN1Lep again demonstrates
1705 improved performance with respect to GN1, in particular at lower p_T . The relative
1706 increase in the b -jet tagging efficiency increases from 4% to 8% with respect to DL1r.
1707 For jets in the Z' sample, GN1 again outperforms DL1r across the entire jet p_T range
1708 studied. The largest relative improvement in performance is found at the highest
1709 transverse momenta of jet $p_T > 2 \text{ TeV}$, and corresponds to an approximate factor of
1710 2 improvement in efficiency with respect to DL1r.

1711 The performance of the model was also evaluated as a function of the average
1712 number of pile-up interactions in the event. No significant dependence of the tagging
1713 performance was observed.

1714 6.5.2 c -tagging Performance

1715 As discussed previously, GN1 does not rely on any inputs from manually optimised
1716 low-level tagging algorithms. Since these algorithms were originally designed and
1717 tuned with the aim of b -tagging, and not c -tagging, the low level tagging algorithms
1718 may perform suboptimally for c -tagging purposes. The tagging of c -jets therefore
1719 presents a compelling use case for GN1. As each of the models is trained with three
1720 output classes, using it as a c -tagging algorithm is trivially analogous to the approach
1721 used for b -tagging. The model output probabilities are combined into a single score
1722 D_c , which is defined similarly to Eq. (6.8) as

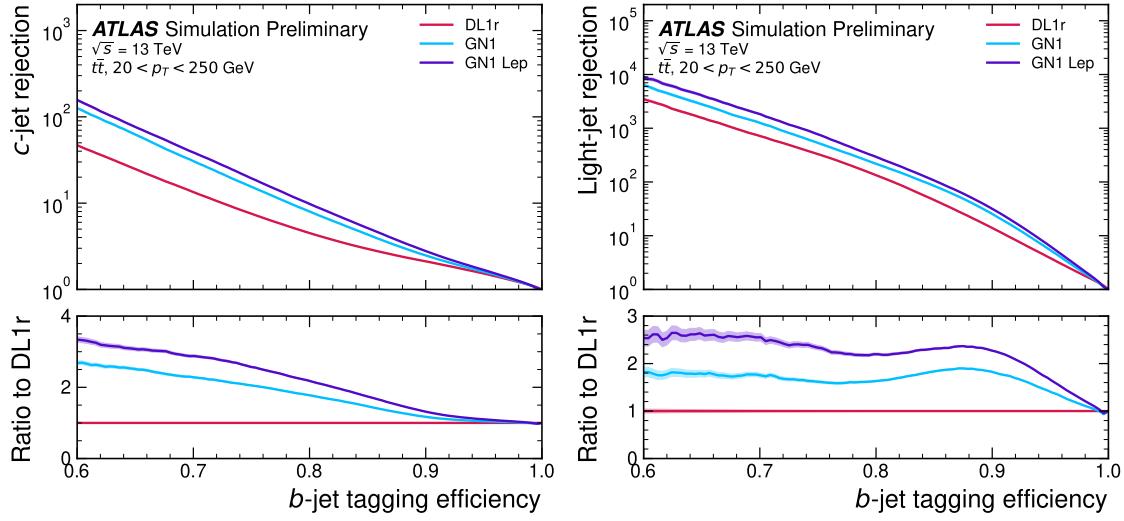


Figure 6.6: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is 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].

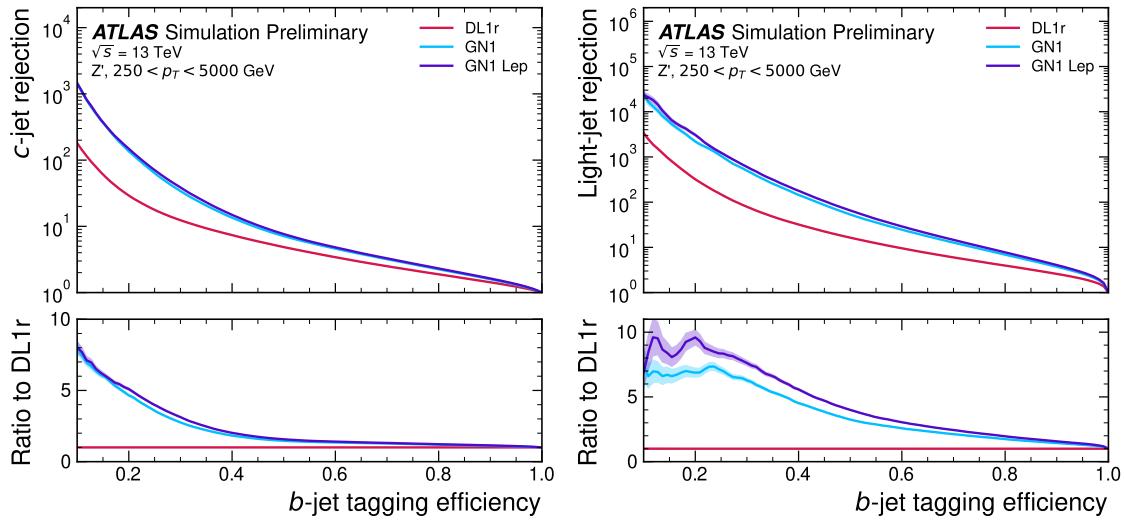


Figure 6.7: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used 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].

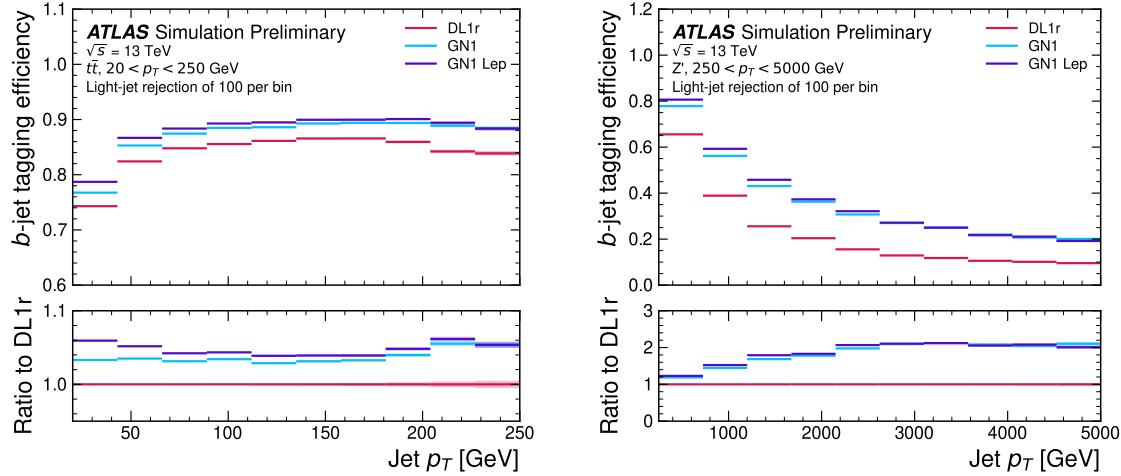


Figure 6.8: The *b*-jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and 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 6.5.1, 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,

1737 with both the b -jet rejection (light-jet rejection) improving with respect to GN1
1738 by approximately 10% (20%) at the 25% c -jet WP. Fig. 6.10 shows the c -tagging
1739 performance on the jets in the Z' sample with $250 < p_T < 5000$ GeV. Both GN1 and
1740 GN1Lep perform similarly, improving the b -jet rejection by 60% and the light-jet
1741 rejection by a factor of 2 at the 25% c -jet WP.

1742 6.5.3 Ablations

1743 Ablation studies (the removal of certain components of a model in order to study
1744 the impact of that component) are carried out to determine the importance of the
1745 auxiliary training objectives of GN1 to the overall performance. The “GN1 No
1746 Aux” variant retains the primary jet classification objective, but removes both track
1747 classification and vertexing auxiliary objectives and correspondingly only minimises
1748 the jet classification loss. The “GN1 TC” variant includes the track classification
1749 objective but not the vertexing objective. Finally, the “GN1 Vert” includes the
1750 vertexing objective, but not the track classification objective.

1751 For jets in both the $t\bar{t}$ and Z' samples, a general trend is observed that the models
1752 trained without one or both of the auxiliary objectives results in significantly reduced
1753 c - and light-jet rejection when compared with the baseline GN1 model, as shown
1754 in Figs. 6.11 and 6.12. For jets in the $t\bar{t}$ sample, the performance of GN1 No Aux
1755 is similar to DL1r, while GN1 TC and GN1 Vert perform similarly to each other.
1756 For jets in the Z' sample, the GN1 No Aux model shows a clear improvement in c -
1757 and light-jet rejection when compared with DL1r at lower b -jet tagging efficiencies.
1758 Similar to jets in the $t\bar{t}$ sample, GN1 TC and GN1 Vert perform similarly, and bring
1759 large gains in background rejection when compared with GN1 No Aux, but the
1760 combination of both auxiliary objectives yields the best performance.

1761 It is notable that the GN1 No Aux model matches or exceeds the performance of
1762 DL1r without the need for inputs from the low-level algorithms. This indicates that
1763 the performance improvements enabled by the improved neural network architecture
1764 used in GN1 appear to be able to compensate for the removal of the low-level
1765 algorithm inputs. The GN1 TC and GN1 Vert variants each similarly outperform
1766 DL1r, demonstrating that both contribute to the overall high performance of the
1767 baseline model. The overall best performing model is the full version of GN1 trained

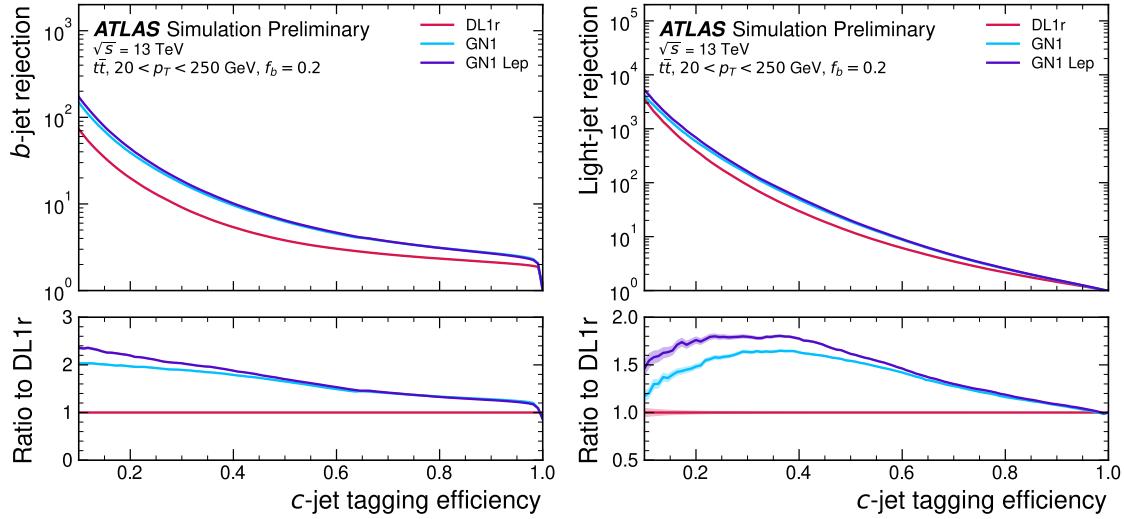


Figure 6.9: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The x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

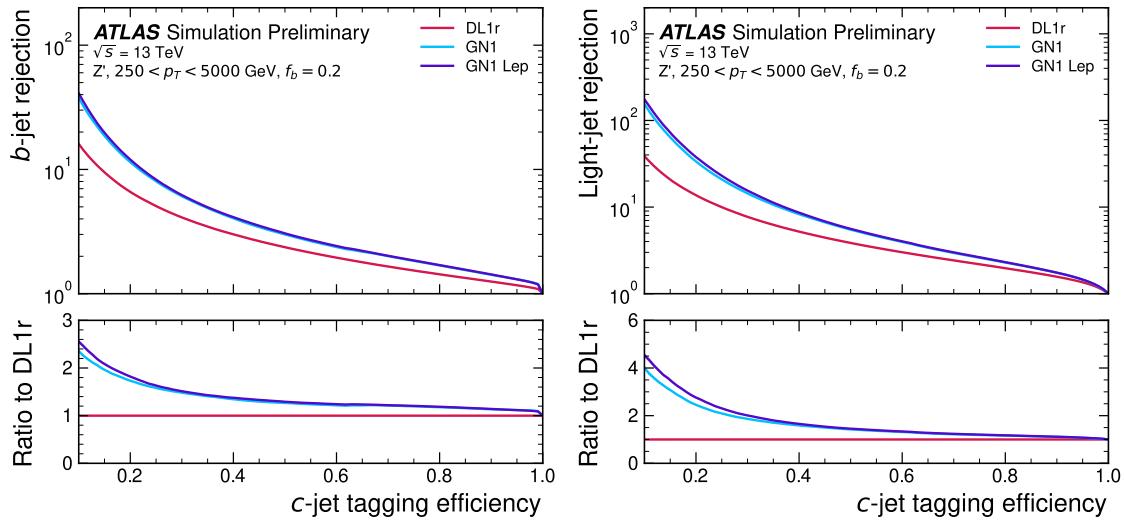


Figure 6.10: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The x -axis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

1768 with both auxiliary objective, demonstrating that the two auxiliary objectives are
1769 complementary.

1770 **6.5.4 Inclusion of Low-Level Vertexing Algorithms**

1771 As already mentioned, GN1 does not include any inputs from the low-level tagging
1772 algorithms, including the vertexing algorithms SV1 and JetFitter [61]. Since these
1773 algorithms are known to play a key role in contributing to the performance of DL1r,
1774 it was studied whether their inclusion in GN1 could result in further performance
1775 improvements. In a dedicated training of GN1, the SV1 and JetFitter tagger outputs
1776 were added to the GN1 jet classification network as an input, similar to how they
1777 are used in DL1r. These outputs include information on the reconstructed vertices,
1778 including the number of vertices, and the properties of the reconstructed vertices.
1779 In addition, if track was used in the reconstruction of a vertex, a corresponding
1780 index to the vertex was included as a track-level inputs to GN1. These indices were
1781 also used to construct an input feature for the edge classification network used to
1782 identify vertices, which was given a value of one if the track-pair were from a common
1783 reconstructed SV1 or JetFitter vertex, and zero otherwise. The jet classification
1784 performance of this GN1 model was not significantly different to the baseline model,
1785 and in some cases the performance was slightly reduced. It was therefore concluded
1786 that GN1 does not benefit from the inclusion of information from SV1 and JetFitter,
1787 indicating that the model is able to reconstruct the relevant information provided
1788 by these low-level algorithms. The study also demonstrates that the model can
1789 function as a highly performant standalone tagger that does not require (beyond
1790 retraining) any manual optimisation to achieve good performance in a wide range
1791 of phase spaces. A dedicated look at the vertexing performance of GN1 with some
1792 comparisons to SV1 and JetFitter is found in Section 6.5.6

1793 **6.5.5 Jet Display Diagrams**

1794 The auxiliary training objectives of GN1 allow for improved model interpretability,
1795 which is especially important for a monolithic approach as the low level taggers,
1796 which provide useful physical insight, are no longer present. Figs. 6.13 and 6.14
1797 provide example comparisons of the true origin and vertexing information compared

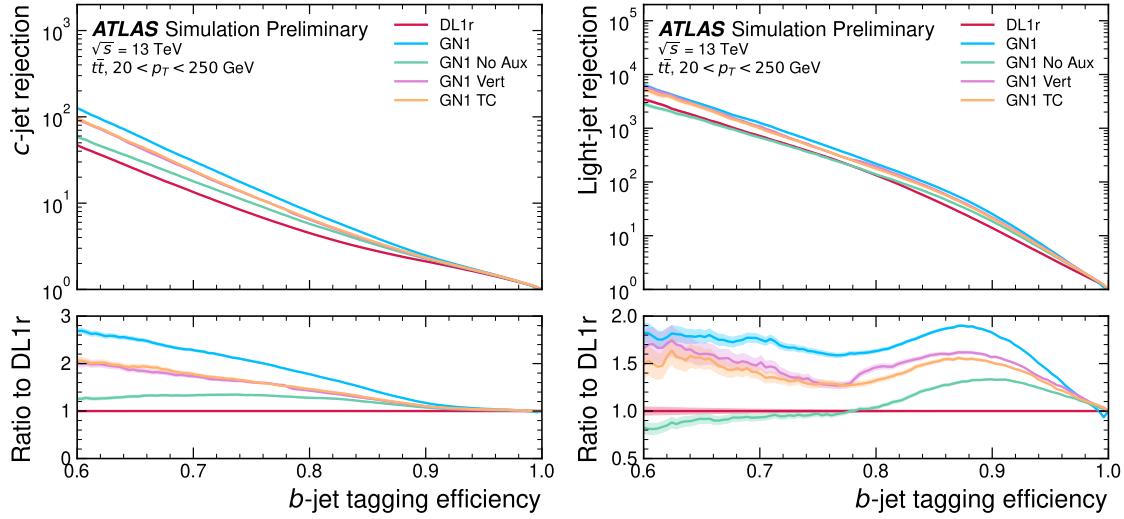


Figure 6.11: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) 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].

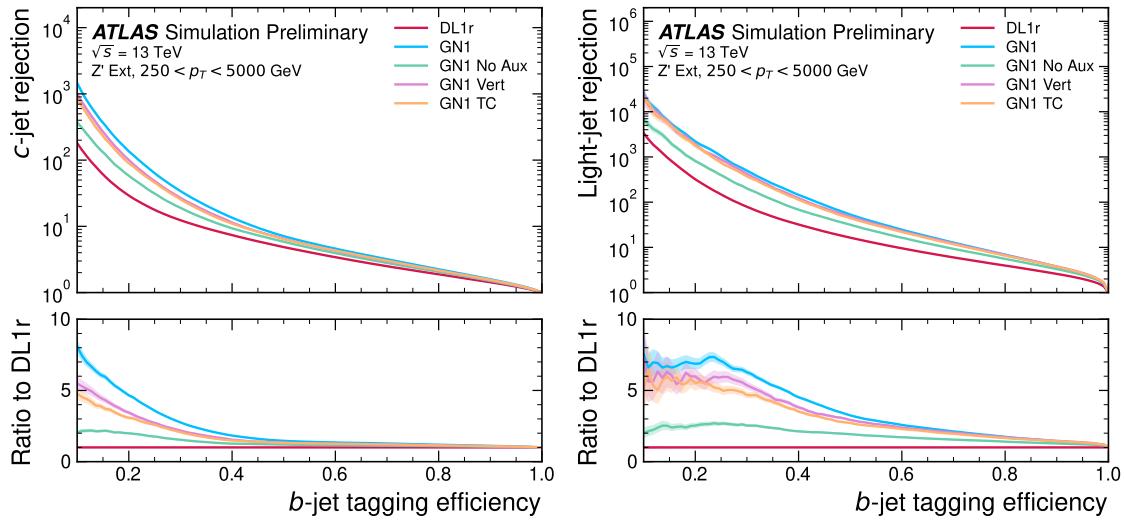


Figure 6.12: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary 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].

1798 with the predicted values from GN1, SV1 and JetFitter. Such comparisons can be
 1799 used to provide an indication that GN1 reconstructs the correct representation of the
 1800 jet structure, and may also help to identify limitations of the model. In the figures,
 1801 the tracks in the jet are indexed twice on each of the x - and y -axes, and tracks are
 1802 grouped into vertices along with other tracks as indicated by common markings in
 1803 the relevant rows and columns.

1804 In Fig. 6.13, GN1 correctly groups the three primary tracks as having come from the
 1805 primary vertex. The b -hadron and $b \rightarrow c$ -hadron decay vertices are also correctly
 1806 predicted, and the origin of the tracks in each is correct. There is a single OtherSec-
 1807 ondary track which GN1 incorrectly predicts as having come from pile-up. Meanwhile
 1808 SV1 (by design) merges the two heavy flavour decay vertices, but incorrectly includes
 1809 a track from the primary vertex. JetFitter reconstructs two vertices, one which is a
 1810 combination of two tracks from different truth vertices and two other single track
 1811 vertices in each of the heavy flavour vertices. GN1 also predicts the flavour of the jet
 1812 with a high degree of certainty.

1813 Similarly Fig. 6.13 shows that GN1 is also able to relatively accurately predict the
 1814 origin and vertex information of tracks inside a jet. The pile-up tracks and primary
 1815 vertex tracks are correctly identified, and the heavy flavour decay tracks are also
 1816 correctly identified with the exception of one of the b -hadron decay tracks. Again,
 1817 SV1 merges the two heavy flavour decay vertices along with a track from pile-up,
 1818 while JetFitter shows signs of being under-constrained by reconstructing two single
 1819 track vertices, one with a pile-up track and one with a track from a $b \rightarrow c$ -hadron
 1820 decay. While these examples do not give a complete picture of the performance of
 1821 GN1, they do show provide a powerful way to visualise and diagnose the behaviour
 1822 of GN1.

1823 6.5.6 Vertexing Performance

1824 From the track-pair vertex prediction, tracks can be partitioned into compatible
 1825 groups representing vertices through the use of a deterministic clustering algorithm
 1826 (see Ref. [115]). As such, GN1 can perform vertex “finding”, but not vertex “fitting”,
 1827 i.e. the reconstruction of a vertex’s properties, which currently still requires the
 1828 use of a dedicated vertex fitter. In order to study the performance of the different
 1829 vertexing tools, the truth vertex label of the tracks are used.

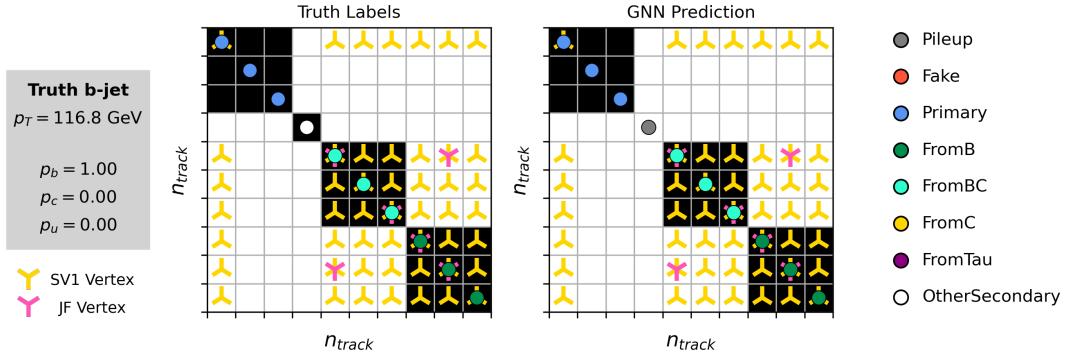


Figure 6.13: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The true and predicted origins of the tracks is shown by the coloured circles along the 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.

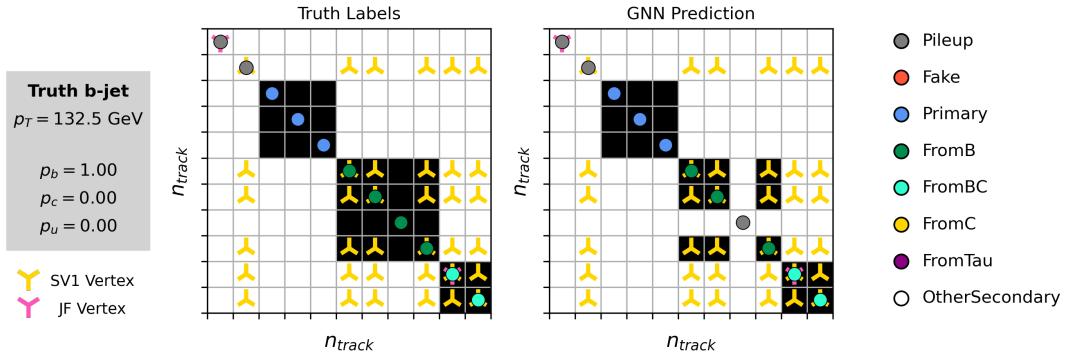


Figure 6.14: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The true and predicted origins of the tracks is shown by the coloured circles along the 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.

1830 There are several caveats to a comparison of the vertexing tools which are a result
1831 of the different approaches they take to vertexing. SV1 and JetFitter are designed
1832 to only find secondary vertices in the jet, whereas GN1 is also trained to determine
1833 which tracks in the jet belong to the primary vertex. To account for this the GN1
1834 vertex with the largest number of predicted primary tracks is excluded from the
1835 vertex finding efficiency calculation. While JetFitter and GN1 aim to resolve each
1836 displaced vertex inside the jet, such that secondary vertices from b -hadron decays
1837 are found separately to tertiary vertices from $b \rightarrow c$ decay chains, SV1 by design
1838 attempts to find a single inclusive vertex per jet. This inclusive vertex groups tracks
1839 from the b -hadron decay itself (FromB) and tracks from $b \rightarrow c$ decays (FromBC). In
1840 order to fairly compare the performance of the different tools, both the exclusive and
1841 inclusive vertex finding efficiencies are studied. For the exclusive vertex finding case
1842 JetFitter and GN1 can be directly compared, while a comparison with SV1 is not
1843 possible due to the aforementioned design constraints. The inclusive vertex finding
1844 performance of all three tools can be compared using the procedure outlined below.

1845 The starting point for the secondary vertex finding efficiency in both the exclusive and
1846 inclusive cases is to select truth HF secondary vertices, defined as those containing
1847 only inclusive b -hadron decays. For exclusive HF vertex finding, these truth secondary
1848 HF vertices can be used directly as the denominator for the efficiency calculation.
1849 Meanwhile for the inclusive efficiency all such truth HF secondary vertices in the jet
1850 are merged into a single inclusive target vertex. Correspondingly, for the inclusive
1851 HF vertex finding case, the vertices found by JetFitter are merged into a single
1852 vertex, and the vertices found by GN1 which contain at least one predicted b -hadron
1853 decay track, are also merged. SV1 does not require any vertex merging. Only jets
1854 containing a single b -hadron at truth level are considered.

1855 Next, vertices in the jet found by the different vertexing tools are compared with
1856 the target truth vertices. The number of correctly and incorrectly assigned tracks is
1857 computed. In order to call a vertex efficient, it is required to contain at least 65% of
1858 the tracks in the corresponding truth vertex, and to have a purity of at least 50%.
1859 Single track vertices are required to have a purity of 100%. Additionally, for GN1
1860 only, at least one track in the vertex is required to have a predicted heavy flavour
1861 origin.

1862 Vertex finding efficiencies for b -jets in the $t\bar{t}$ sample are displayed as a function of
1863 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 flavour vertices found in heavy flavour jets, the improved reconstruction of such vertices may help to differentiate between different flavour of jet.

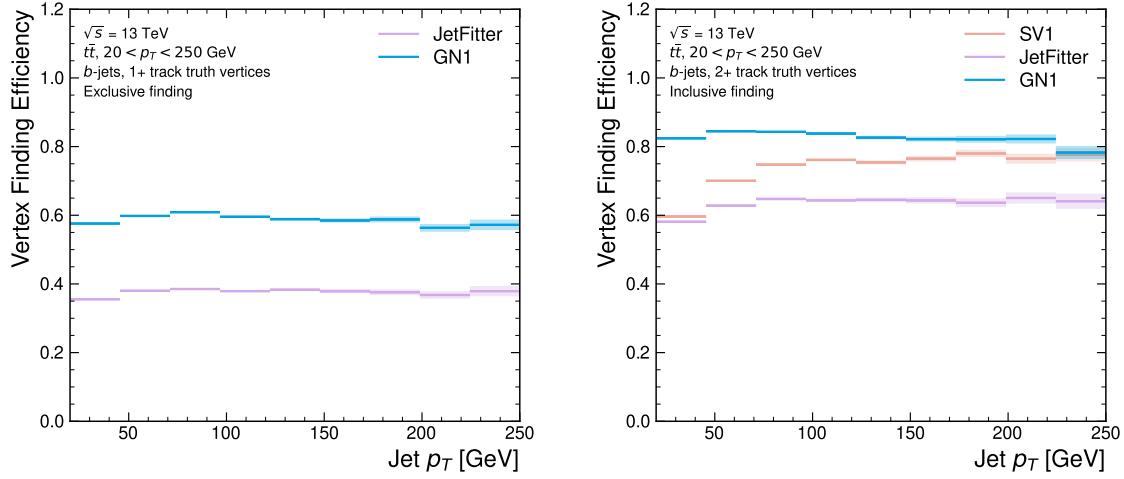


Figure 6.15: Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices 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.

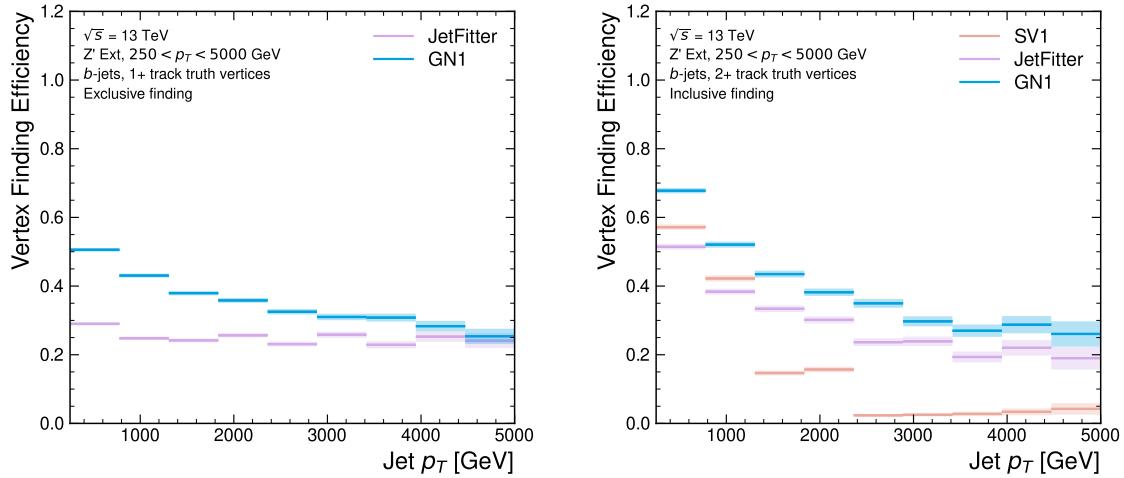


Figure 6.16: Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in the Z' sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain 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.

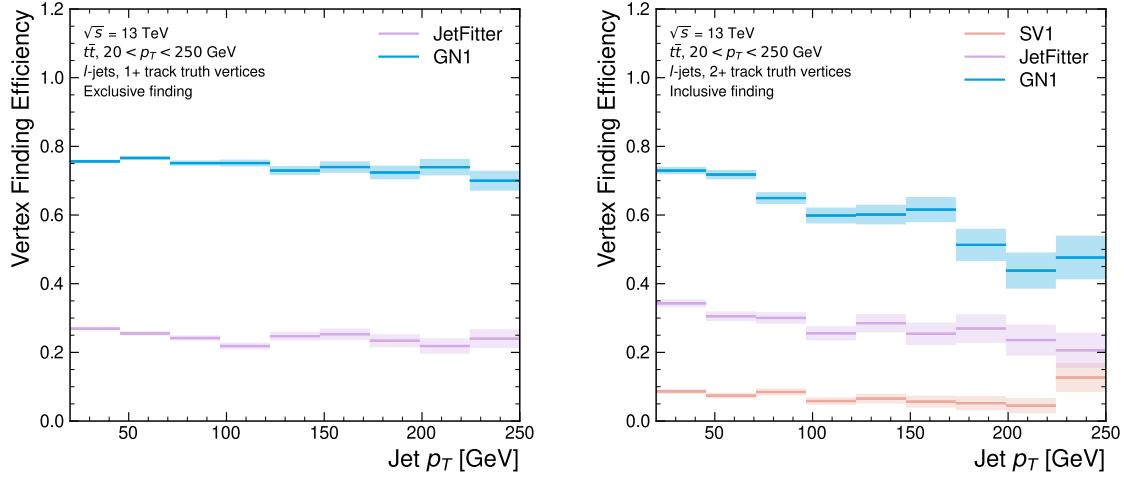


Figure 6.17: Vertex finding efficiency for other secondary decays as a function of jet p_T for light-jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires 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.

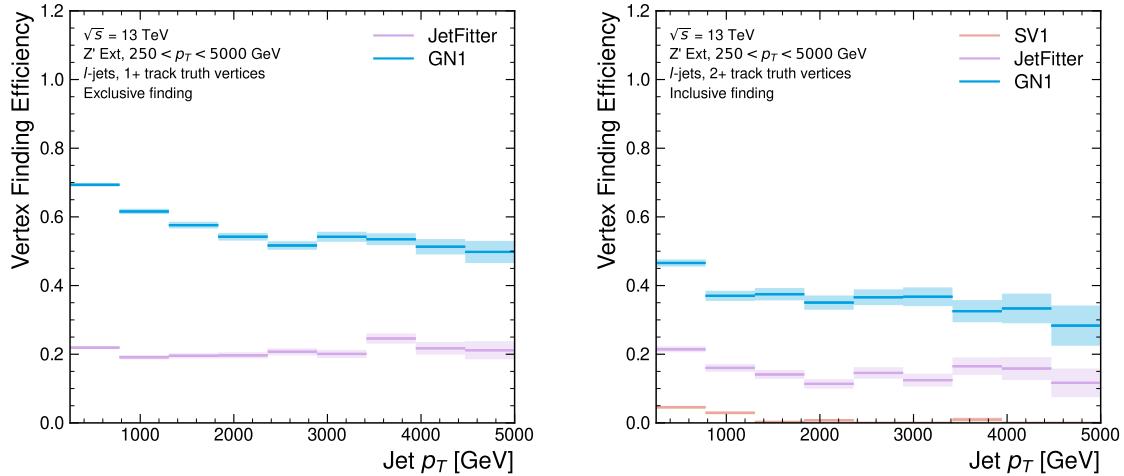


Figure 6.18: Vertex finding efficiency as a function of jet p_T for light-jets in the Z' sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least 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.

1898 Collectively, the results in this section demonstrate that GN1 is able to accurately
1899 group tracks by their spatial origin in both b -jets and light-jets. The purity of the
1900 found vertices was also investigated and was found to be comparable or better than
1901 that of SV1 and JetFitter.

1902 6.5.7 Track Classification Performance

1903 One of the two auxiliary training objectives used by GN1 is to predict the truth origin
1904 of each track associated to the jet. Since the equivalent information is not provided
1905 by any of the existing flavour tagging tools, a benchmark model used to predict the
1906 truth origin of each track is trained based on a standard multi-class feed-forward
1907 classification network. The benchmark model is trained on the same tracks used for
1908 the baseline GN1 training. The model uses precisely the same concatenated track-
1909 and-jet inputs as used by GN1, but processes only a single track at a time, meaning
1910 it cannot take into account the correlations between tracks when determining the
1911 track origin. The model is made up of five densely connected linear layers with 200
1912 neurons in each layer. The performance of the model was found to be insensitive to
1913 changes in the network structure.

1914 To measure the track classification performance, the area under the curve (AUC) of
1915 the receiver operating characteristic (ROC) curve is computed for each origin class,
1916 using a one-versus-all classification approach². The AUCs for the different truth
1917 origins are averaged using both an unweighted and a weighted mean. The unweighted
1918 mean treats the performance of each class equally, while the weighted mean uses
1919 as a weight the relative abundance of tracks of each class. The same approach is
1920 used to compute the precision, recall and F1 scores³. Table 6.3 demonstrates clearly
1921 that GN1 outperforms the MLP both at $20 < p_T < 250$ GeV for jets in the $t\bar{t}$ sample
1922 and at $250 < p_T < 5000$ GeV for jets in the Z' sample. For example, GN1 can reject
1923 65% of fake tracks in jets in the $t\bar{t}$ sample, while retaining more than 99% of good
1924 tracks (i.e. those tracks which are not fake). The GN1 model has two advantages
1925 over the MLP which can explain the performance improvement. Firstly, the graph

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

³The 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.

1926 neural network architecture enables the sharing of information between tracks. This
1927 is likely to be beneficial since the origins of different tracks within a jet are correlated.
1928 Secondly, the jet classification and vertexing objectives may be complementary to the
1929 track classification objective, and so the track classification performance is improved
1930 by the combined training of complementary objectives.

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

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.

1931 Fig. 6.19 shows the track origin classification ROC curves for the different track
1932 origins for jets in both the $t\bar{t}$ and Z' samples. In order to improve visual readability
1933 of the plot, the curves for the heavy flavour truth origins (FromB, FromBC and
1934 FromC) have been combined (weighted by their relative abundance), as have the
1935 Primary and OtherSecondary origins. In jets in both the $t\bar{t}$ and Z' samples, the AUC
1936 of all the different origin groups exceeds 0.9, representing strong overall classification
1937 performance. In both samples fake tracks are the easiest to classify, followed by
1938 pile-up tracks. The FromC tracks which are c -hadron decay products, are the hardest
1939 to classify, possibly due to their similarity to both fragmentation tracks and b -hadron
1940 decay tracks, depending on the c -hadron species in question.

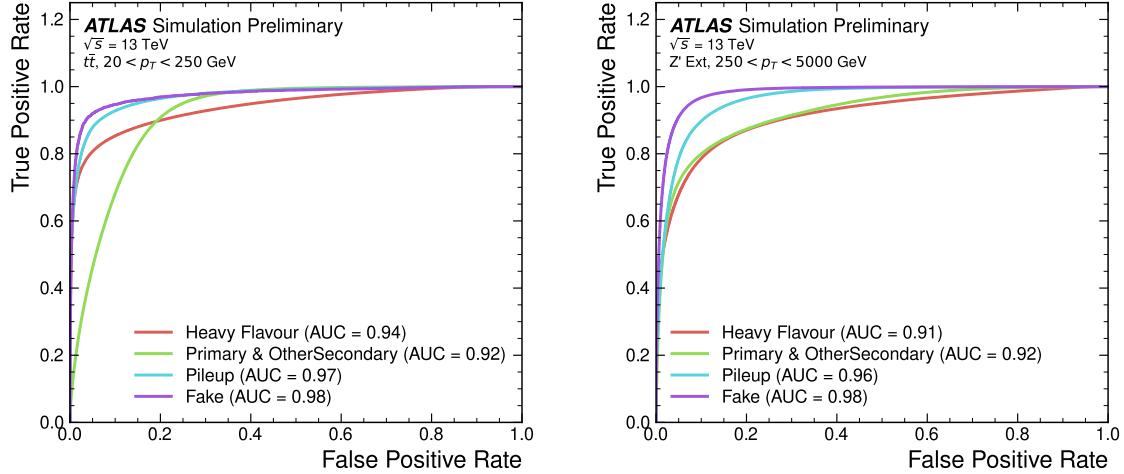


Figure 6.19: ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right). The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the 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].

1941 6.6 Further Studies

1942 6.6.1 Looser Track Selection

1943 The track selections used to produce the main results are listed in Table 3.2. This
 1944 includes a selection on the number of shared silicon modules used to reconstruct the
 1945 track $N_{\text{shared}}^{\text{Si}}$. This value is calculated as

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

1946 where $N_{\text{shared}}^{\text{Pix}}$ is the number of shared pixel hits and $N_{\text{shared}}^{\text{SCT}}$ is the number of shared
 1947 SCT modules on a track. The nominal selection used elsewhere in this thesis is
 1948 $N_{\text{shared}}^{\text{Si}} < 2$. As the rate of shared hits is significantly higher for b -hadron decay
 1949 tracks than for other tracks, especially at high- p_T , this selection rejects a significant
 1950 proportion of these tracks.

1951 Figs. 6.20 and 6.21 show the result of training the GN1 tagger with the full relaxation
 1952 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.

6.6.2 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 tracks⁴ 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 [90], and two versions of DIPS [87], 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.

6.6.3 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. [126]. Figs. 6.23 and 6.24 are reproduced from Ref. [126]. 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

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

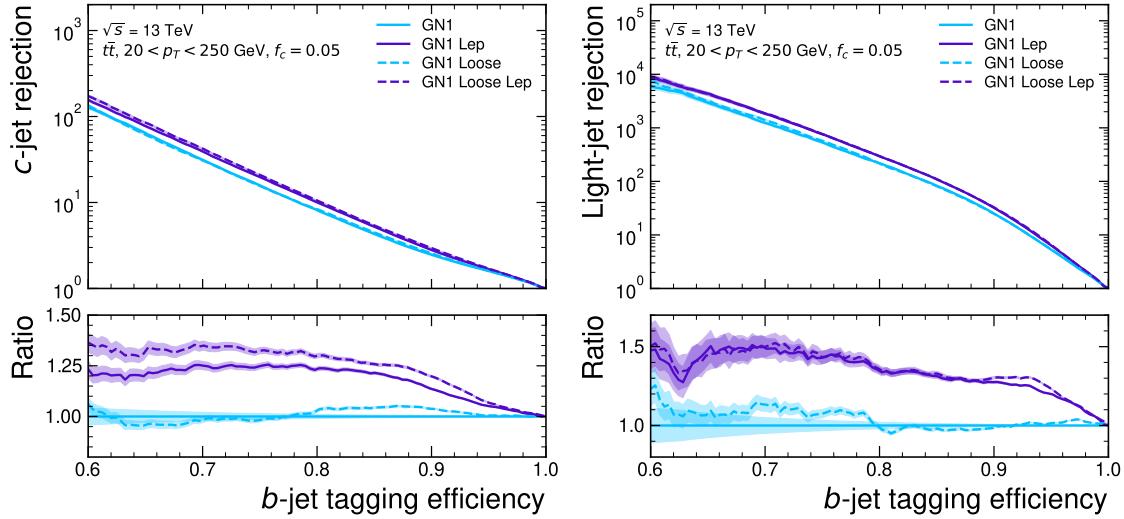


Figure 6.20: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the baseline GN1 model is shown in the bottom panels. A value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

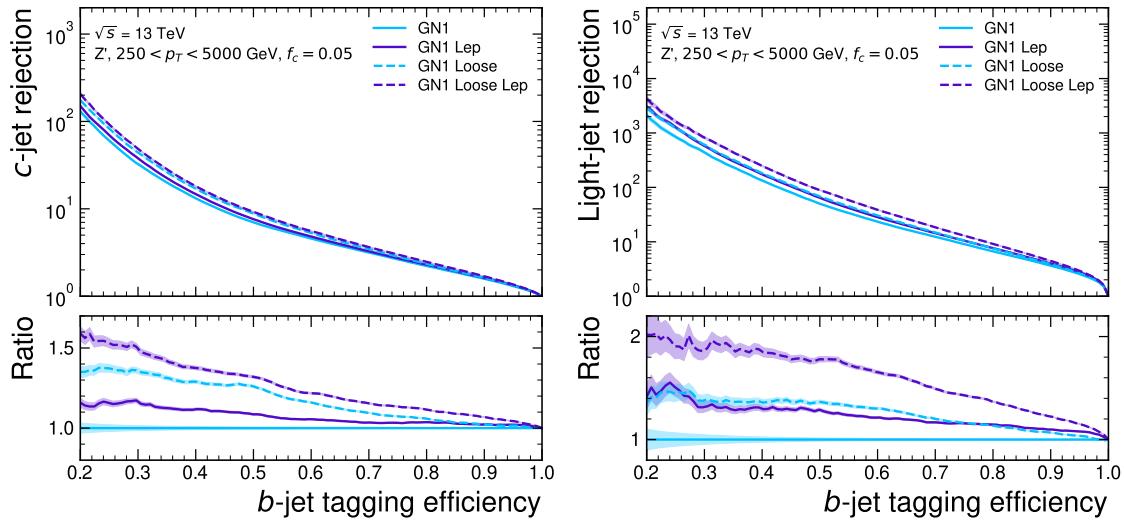


Figure 6.21: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the baseline GN1 model is shown in the bottom panels. A value of $f_c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

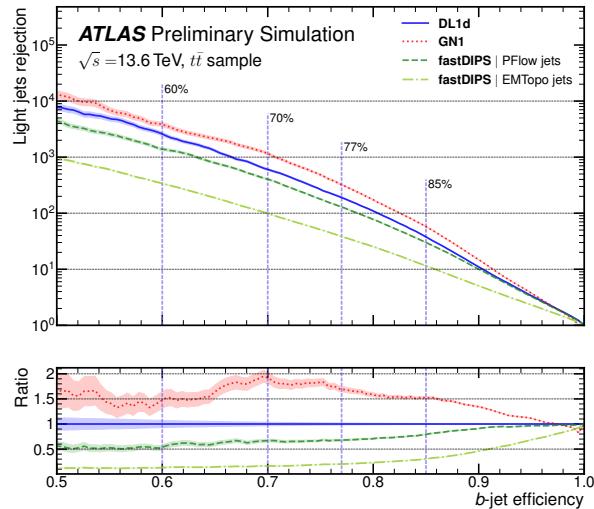


Figure 6.22: The HLT light-jet rejection as a function of the b -jet efficiency jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ for events with a centre of mass energy $\sqrt{s} = 13.6 \text{ TeV}$ [125]. The ratio to the performance of the DL1d algorithm [90] is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for b-tagging.

1981 DL1d [90], GN1 improves on the c -jet rejection (light-jet rejection) by a factor of ~ 2
1982 (~ 2.5) for jets in the $t\bar{t}$ sample at the 60% b -jet WP. Significant improvements in
1983 rejections are also observed for jets in the Z' sample.

1984 6.7 Conclusion

1985 In this chapter a novel jet tagger, GN1, is presented. The model has a graph neural
1986 network architecture and is trained with auxiliary training objectives, which are
1987 shown to improve the performance of the basic model. GN1 significantly improves
1988 flavour tagging performance with respect to DL1r, the current default ATLAS flavour
1989 tagging algorithm, when compared in simulated collisions. GN1 improves c - and
1990 light-jet rejection for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV by factors of
1991 ~ 2.1 and ~ 1.8 respectively at a b -jet tagging efficiency of 70% when compared with
1992 DL1r. For jets in the Z' sample with $250 < p_T < 5000$ GeV, GN1 improves the
1993 c -jet rejection by a factor of ~ 2.8 and light-jet rejection by a factor of ~ 6 for a
1994 comparative b -jet efficiency of 30%.

1995 Previous multivariate flavour tagging algorithms relied on inputs from low-level
1996 tagging algorithms, whereas GN1 needs no such inputs, making it more flexible.
1997 It can be easily fully optimised via a retraining for specific flavour tagging use
1998 cases, as demonstrated with c -tagging and high- p_T b -tagging, without the need for
1999 time-consuming retuning of the low-level tagging algorithms. The model is also
2000 simpler to maintain and study due to the reduction in the number of constituent
2001 components.

2002 GN1 demonstrates improved track classification performance when compared with a
2003 simple per-track MLP. The model is also able to perform vertex finding, and prelimi-
2004 nary studies suggest it outperforms previous manually optimised approaches. The
2005 auxiliary track classification and vertex finding objectives are shown to significantly
2006 contribute to the performance in the jet classification objective, and, along with the
2007 more advanced graph neural network architecture, are directly responsible for the
2008 improvement over DL1r.

2009 GN1 has also been shown to perform well for b -tagging in the High Level Trigger,
2010 and in the projected environment for the upgraded High Luminosity LHC. The

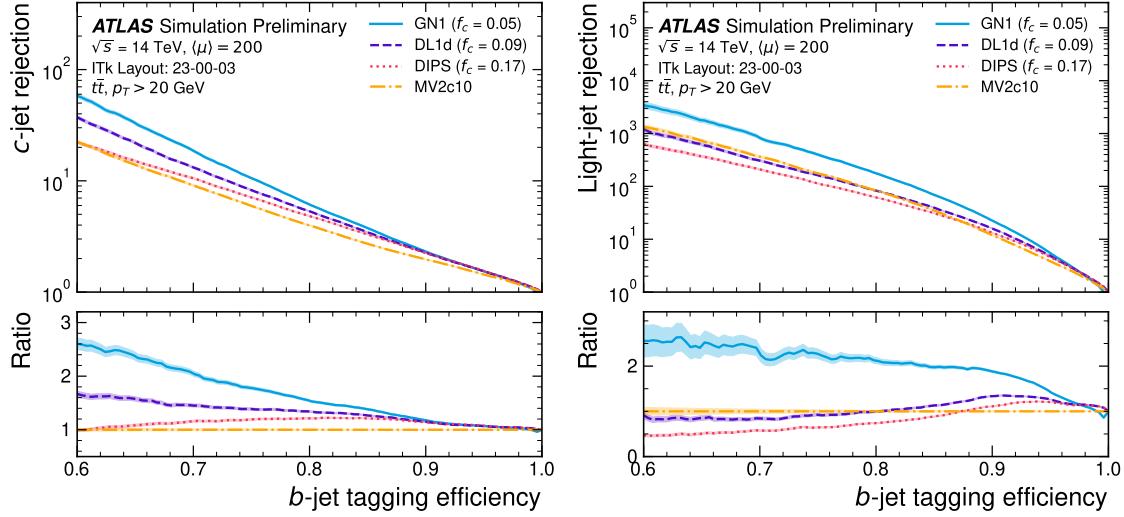


Figure 6.23: The c -jet rejection (left) and light-jet rejection (right) for the upgraded HL-LHC ATLAS detector with $\langle \mu \rangle = 200$ as a function of the b -jet efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of mass energy $\sqrt{s} = 14$ TeV [126]. 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.

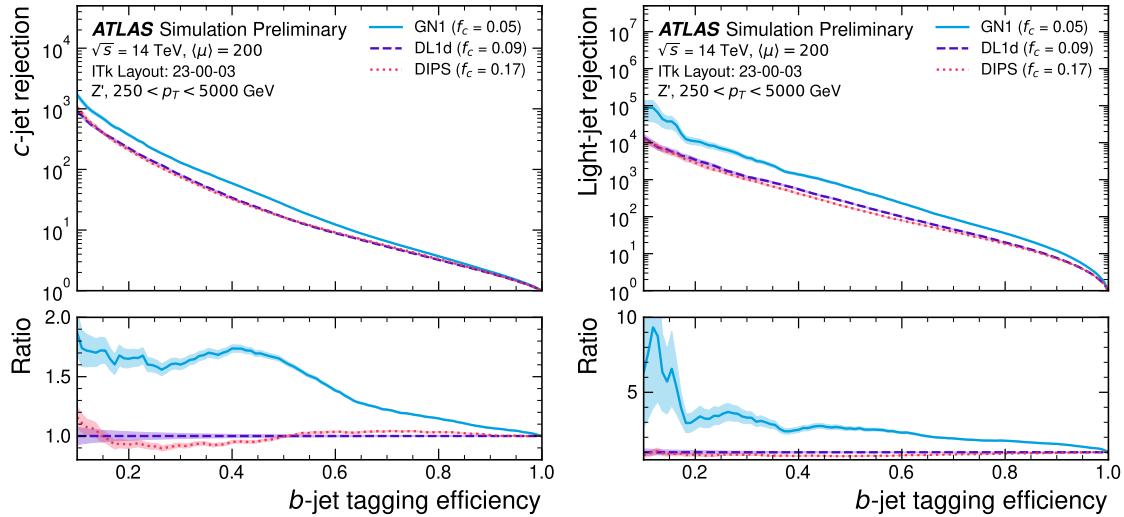


Figure 6.24: The c -jet rejection (left) and light-jet rejection (right) for the upgraded HL-LHC ATLAS detector with $\langle \mu \rangle = 200$ as a function of the b -jet efficiency for jets in the Z' sample with $250 < p_T < 5000$ GeV for events with a centre of mass energy $\sqrt{s} = 14$ TeV [126]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions.

2011 performance of the model at high- p_T can be improved by loosening the input track
2012 selection, which increases the number of b -hadron decay tracks that are inputted to
2013 the model.

2014 Preliminary validation of the model demonstrates that the level of discrepancy
2015 between different Monte Carlo event generators is similar to that seen for previous
2016 ATLAS ML based taggers such as DL1r, suggesting that the model has not learnt
2017 additional information that is specific to a single event generator. Initial comparisons
2018 between simulated events and data also show similar levels of agreement as for
2019 previous taggers [127].

2020 **Future Work**

2021 Further improvements in the b - and c -tagging performance are likely possible with
2022 a more thorough optimisation of the model architecture, and the integration of
2023 additional information from other parts of the ATLAS detector. The addition of
2024 other auxiliary training objectives, such as the prediction of the truth b -hadron decay
2025 radius and transverse momentum and the truth type of input tracks may also yield
2026 additional performance gains.

2027 Although the results demonstrate a significant performance improvement at high- p_T ,
2028 it is also possible that additional studies on further loosening the selection could yield
2029 further improved results. For example the selections on the number of number of
2030 holes and the longitudinal impact parameter could be further relaxed. The maximum
2031 number of tracks provided as input to the model could also be increased from the
2032 default value of 40. In order to validate the change from the default tracking setup,
2033 studies investigating the modelling uncertainties of the additional tracks need to be
2034 carried out.

2035 Given that GN1 exploits the input information in a more sophisticated way than
2036 DL1r, further studies are needed to confirm that the performance gain observed in
2037 these simulated samples is also observed in experimental data. Additional future
2038 work includes the full calibration of the model so it can be used by physics analyses.

2039 The flexible nature of the model means it can also be readily applied to other related
2040 problems outside of standard b - and c -tagging applications, as demonstrated in
2041 Section 6.6. Additional applications for the architecture include $X \rightarrow bb$ and $X \rightarrow cc$

2042 tagging. The model could also be repurposed as a pile-up jet or τ tagger, or general
2043 primary and secondary vertexing tool. Using a vertex fitting algorithm to compare
2044 the reconstructed vertex quantities with those from SV1 and JetFitter is left for
2045 future work.

2046 Chapter 7

2047 Boosted VHbb Analysis

2048 The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [14, 15],
2049 is predicted by the Standard Model to decay primarily to a pair of b -quarks, with
2050 a branching fraction of 0.582 ± 0.007 for $m_H = 125$ GeV [26]. Observation of this
2051 decay mode was reported by ATLAS [79] and CMS [28] in 2018, establishing the
2052 first direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks
2053 (see Section 2.2.2). The $H \rightarrow b\bar{b}$ process is also important for constraining the total
2054 decay width of the Higgs [128].

2055 Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion
2056 as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet
2057 background and so overall sensitivity to the Higgs is low. The QCD multijet
2058 background refers to events containing one or more strongly produced jets which
2059 are not the decay product of heavy resonances, for example $g \rightarrow q\bar{q}$. In the $H \rightarrow b\bar{b}$
2060 gluon-gluon fusion channel the vast majority of events will contain only jets in the
2061 final state, and therefore it is extremely difficult to distinguish signal events from the
2062 overwhelming multijet background. The $H \rightarrow b\bar{b}$ observation therefore searched for
2063 Higgs bosons produced in association with a vector boson V (where V can be a W
2064 or Z boson) which subsequently decays leptonically. The leptonic final states allow
2065 for leptonic triggering whilst at the same time significantly reducing the multijet
2066 background.

2067 Two full Run 2 dataset analyses were carried out as a follow-up to the $H \rightarrow b\bar{b}$
2068 observation [79]. Similar to the observation, both measured the associated production
2069 of a Higgs with a vector boson, with the Higgs boson decaying to a pair of b -quarks.
2070 The first analysis [129] was focussed on the resolved phase-space, where the Higgs

2071 candidate is reconstructed as two distinct jets with radius parameter $R = 0.4$.
2072 The second analysis [130] was focussed on the boosted phase-space, where the
2073 Higgs candidate had a sufficiently large transverse momenta such that it can be
2074 reconstructed as a single jet with a radius parameter of $R = 1.0$. This chapter
2075 will focus on the latter analysis, which is referred to as the boosted VH , $H \rightarrow b\bar{b}$
2076 analysis.

2077 In this chapter, the boosted VH , $H \rightarrow b\bar{b}$ analysis is outlined in Section 7.1. An
2078 introduction to the systematic uncertainties used in the analysis is provided in
2079 Section 7.1.6. Detailed information about the modelling studies performed for
2080 the analysis are described in Section 7.3. The statistical treatment is detailed in
2081 Section 7.4, and the results of the analysis are presented in Section 7.5. This analysis
2082 has been published in Ref. [130]. Figures and tables from the published work are
2083 reproduced here.

2084 7.1 Analysis Overview

2085 The boosted VH , $H \rightarrow b\bar{b}$ analysis is focused on the high transverse momentum
2086 regime, which has the benefit of being more sensitive to physics beyond the Standard
2087 Model [131], but the disadvantage of being more challenging due to the increased
2088 difficulty in the accurate reconstruction of high transverse momentum physics objects
2089 (discussed in Chapter 4). In order to focus on the high- p_T regime, the reconstructed
2090 vector boson p_T^V is required to be $p_T^V > 250$ GeV (see Section 7.1.1). Events are
2091 also split into two p_T bins with the first bin covering $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and
2092 the second covering $p_T^V > 400$ GeV, which allows the analysis to benefit from the
2093 improved signal-to-background in the high- p_T regime.

2094 The previous ATLAS analysis in Ref. [79] was primarily sensitive to vector bosons
2095 with a more modest p_T^V boost in the region of 150–300 GeV (for the 0- and 1-lepton
2096 channels) and 75–300 GeV (for the 2-lepton channel). In this regime, the Higgs
2097 candidate was reconstructed using a pair of jets with radius parameter of $R = 0.4$,
2098 called small- R jets. However in the high- p_T regime, the decay products of the Higgs
2099 boson become increasingly collimated and the small- R jets may not be individually
2100 resolved. In order to enhance the reconstruction of the Higgs boson candidate, this
2101 analysis uses a large- R jet with radius parameter $R = 1.0$ to reconstruct the Higgs

2102 boson candidate. The Higgs candidate is required to have exactly two ghost-associated
2103 and b -tagged variable-radius track-jets (b -track-jets). The candidate large- R jet is
2104 reconstructed using jet substructure techniques, in particular it is trimmed by
2105 removing soft and wide-angle components, which helps to remove particles from the
2106 underlying event and pile-up collisions [132]. Refer to Section 3.4.3 for more details
2107 on jet reconstruction.

2108 On top of the binning in p_T^V , selected events are further categorised into the 0-, 1-
2109 and 2-lepton channels depending on the number of charged leptons (electrons and
2110 muons) present in the reconstructed final state (also referred to as the 0L, 1L, and 2L
2111 channels respectively). The 0-lepton channel targets the $ZH \rightarrow \nu\nu b\bar{b}$ process, the
2112 1-lepton channel targets $WH \rightarrow \ell\nu b\bar{b}$, and the 2-lepton channel targets $ZH \rightarrow \ell\ell b\bar{b}$,
2113 where ℓ is an electron or muon and ν is a neutrino. Each channel has a dedicated
2114 set of selections which are listed in more detail in Section 7.1.2. Events in the 0- and
2115 1-lepton channels are further split depending on the number of additional small- R
2116 jets not matched to the Higgs candidate. The high-purity signal region (HP SR)
2117 has zero such jets, while the low-purity signal region (LP SR) has one or more, and
2118 therefore absorbs a larger number of background $t\bar{t}$ events. Maintaining a high purity
2119 signal region is important for the extraction of the signal yield. The 0- and 1-lepton
2120 channels also make use of a dedicated $t\bar{t}$ control region for jets with one or more
2121 additional b -tagged small- R jets, described in Section 7.1.3. A complete overview of
2122 the different analysis regions is given in Table 7.1.

2123 The large- R jet mass is used as the main discriminant in the analysis. The signal
2124 $VH, H \rightarrow b\bar{b}$ yields is extracted from a profile likelihood fit to the large- R jet mass
2125 over several signal and control analysis regions, which are described in Sections 7.1.2
2126 and 7.1.3. The diboson background $VZ, Z \rightarrow b\bar{b}$ yield is simultaneously extracted
2127 from the fit, and provides a cross check on the signal extraction. The fit model
2128 (described henceforth only as “the fit”) is described in more detail in described in
2129 Section 7.4.

2130 7.1.1 Object Reconstruction

2131 The presence of neutrinos in the $WH \rightarrow \ell\nu b\bar{b}$ and $ZH \rightarrow \ell\ell b\bar{b}$ signatures can be
2132 inferred from a momentum imbalance in the transverse plane (see Section 3.4.5). The

		Analysis Regions					
		250 < p_T^V < 400 GeV			$p_T^V \geq 400$ GeV		
Channel	0 add. b -track-jets		≥ 1 add. b -track-jets	0 add. b -track-jets		≥ 1 add. b -track-jets	
	0 add. small- R jets	≥ 1 add. small- R jets		0 add. small- R jets	≥ 1 add. small- R jets		
0-lepton	HP SR	LP SR	CR	HP SR	LP SR	CR	
1-lepton	HP SR	LP SR	CR	HP SR	LP 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 [130].

2133 vector boson transverse momentum p_T^V is reconstructed as the missing transverse
 2134 energy E_T^{miss} in the 0-lepton channel, as the magnitude of the summed $\mathbf{E}_T^{\text{miss}}$ and
 2135 charged-lepton momentum in the 1-lepton channel, and as the transverse momentum
 2136 of the 2-lepton system in the 2-lepton channel.

2137 Electrons and muons are reconstructed and identified as outlined in Section 3.4.4,
 2138 and following the approach described in Ref. [79]. Leptons are required to satisfy
 2139 the selections listed in Table 7.2. *Baseline* electrons are required to pass the looser
 2140 likelihood-based identification selection described in Section 3.4.4, whilst *Signal*
 2141 electrons are required to satisfy a tighter likelihood identification selection. *Baseline*
 2142 muons are required to pass the ‘loose’ identification described in Ref. [51], while
 2143 *signal* muons are required to pass the ‘medium’ identification working point. All
 2144 signal leptons are required to additionally satisfy a $p_T > 27$ GeV selection criteria,
 2145 except for muons in the 1-lepton channel where a cut of 25 GeV is used. The number
 2146 of baseline leptons is used to categorise the event into the 0-, 1- or 2-lepton channels.
 2147 The 1- and 2-lepton channels additionally require one signal lepton to be present.

2148 The analysis makes use of large- R and variable-radius small- R track-jets, which are
 2149 described in Section 3.4.3. The large- R jets are required to satisfy $p_T > 250$ GeV and
 2150 $|\eta| < 2.0$, and are used to reconstruct the Higgs boson candidate, while small- R jets
 2151 are used for b -tagging and for selection of the analysis region. These track-jets are

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 GeV in the 1-lepton channel).

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 [133–135]. The jet tagging strategy relies on extensive studies into track-jet b -tagging in boosted topologies [136, 137].

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^{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^{miss} trigger as the 0-lepton channel is used. Since muons are not used for the E_T^{miss} trigger calculations, this is in effect a p_T requirement on

2174 the muon-neutrino system, which in the analysis phase space is more efficient than a
2175 single-muon trigger. The 2-lepton channel uses the same triggering strategy as the
2176 1-lepton channel. In all channels, the trigger selections applied are fully efficient for
2177 events selected using the full requirements in Table 7.3.

2178 The combined selections in Table 7.3 result in a signal efficiency ranging from 6–16%
2179 for the WH and ZH processes depending on the channel and p_T^V bin.

2180 7.1.3 Control Regions

2181 The $t\bar{t}$ process presents a major background in the 0- and 1-lepton channels. In these
2182 events, the Higgs candidate is often reconstructed from a correctly tagged b -jet from
2183 the top decay $t \rightarrow Wb$, and an incorrectly tagged c - or light-jet from the subsequent
2184 decay of the W , as shown in Fig. 7.1.

2185 The top quark decays to a W and a b -quark the vast majority of the time. Hence,
2186 the second top quark from the $t\bar{t}$ pair is also likely to result in a second tagged
2187 b -tagged track-jet outside of the large- R Higgs candidate. To ensure sufficient $t\bar{t}$
2188 rejection, 0- and 1-lepton channel signal regions are defined using a veto on events
2189 with b -tagged track-jets outside the Higgs candidate. These events are used to
2190 construct a control region (CR) which is enriched in $t\bar{t}$ events. The CR is used to
2191 constrain the normalisation of the $t\bar{t}$ background in the fit.

2192 7.1.4 Background Composition

2193 After the selections described in Section 7.1.2 the number of background events
2194 mimicking the VH , $H \rightarrow b\bar{b}$ signal is greatly reduced. However, the number of
2195 background events still greatly outnumbers that of signal events. The background
2196 processes are channel dependent. In the 0-lepton channel the dominant sources of
2197 backgrounds are Z +jets ($Z \rightarrow \nu\nu$) and $t\bar{t}$, with W +jets and diboson events being
2198 subdominant. W +jets contribute in the event of $W \rightarrow \tau\nu$, and subsequent hadronic
2199 decay of the τ or lack of successful reconstruction/selection of the leptonic decay
2200 products. $t\bar{t}$ and W +jets (with a leptonic decay of the W as in $W \rightarrow \ell\nu$) are
2201 dominant in the 1-lepton channel, while single-top is subdominant. In the 2-lepton
2202 channel, Z +jets ($Z \rightarrow \ell\ell$) is again dominant followed by ZZ diboson events.

Selection	0 lepton channel	1 lepton channel	2 leptons channel
Trigger	E_T^{miss}	e sub-channel μ sub-channel	e sub-channel μ sub-channel
Leptons	0 <i>baseline</i> leptons	Single electron 1 <i>signal</i> lepton $p_T > 27 \text{ GeV}$	Single electron 2 <i>baseline</i> leptons among which $p_T > 25 \text{ GeV}$ no second <i>baseline</i> lepton ≥ 1 <i>signal</i> lepton, $p_T > 27 \text{ GeV}$ both leptons of the same flavour -
E_T^{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	
<i>b</i> -tagged jets		leading two track-jets matched to the leading large- R must be <i>b</i> -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 [130]. 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_{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.

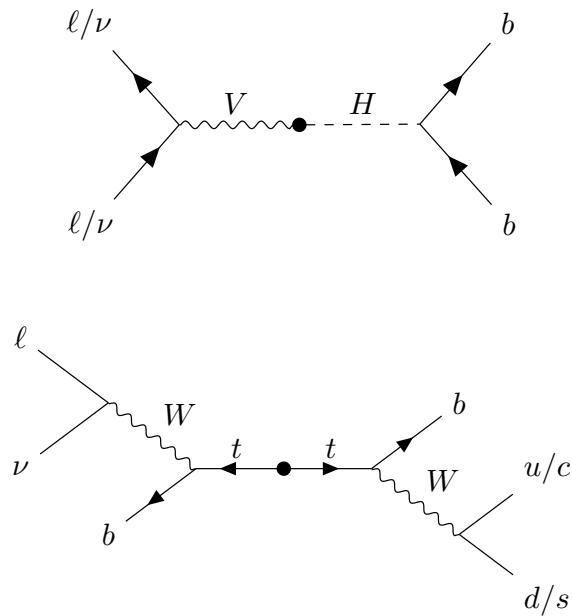


Figure 7.1: Diagrams of the signal process (top) and $t\bar{t}$ background (bottom). The object to the right of centre are reconstructed within the large- R jet. For the $t\bar{t}$ background, the large- R jet contains a mistagged c -jet or (less often) a mistagged light-jet. The contribution in the 0-lepton channel results from hadronically decaying τ 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.

- 2203 The diboson background VV consists primarily of WZ and ZZ events in which the
2204 Z decays to a pair of b -quarks. This process very closely matches the signal, with a
2205 resonant peak occurring at $m_Z = 91\text{ GeV}$. As it is a well understood process, the
2206 diboson background can be used as a cross-check on the Higgs results.
- 2207 The $t\bar{t}V$, $t\bar{t}H$ and multijet backgrounds are negligible in the analysis phase space
2208 after the selections have been applied, with the exception of the 1-lepton electron
2209 sub-channel, in which multijet background is not negligible. The multijet background
2210 in this region is generally made up of events where the isolated leptonic signature
2211 has been mimicked by either a jet or electron from a semi-leptonic heavy flavour
2212 decay, where the electron has escaped the jet.
- 2213 The contributions from the different backgrounds are modelled using Monte Carlo
2214 event generators and the uncertainties associated with these samples are studied in
2215 Section 7.3. The multijet background is modelled using a data-driven technique.

2216 7.1.5 Data & Simulated Samples

- 2217 Data from centre-of-mass energy $\sqrt{s} = 13\text{ TeV}$ proton-proton collisions at the LHC
2218 recorded over the course of Run 2 (between 2015 and 2018) were used for the analysis.
2219 The resulting dataset corresponds to a total integrated luminosity of 139 fb^{-1} .
- 2220 An overview of the MC simulated samples used in the analysis is given in Table 7.4.
2221 These samples are used to model the signal and background processes relevant to the
2222 analysis, with the exception of the multijet background which is modelled using a
2223 data-driven technique. Data and simulated events are reconstructed using the same
2224 algorithms, and a reweighting is applied to the simulated events in order to match
2225 the pile-up distribution observed in the data.

2226 7.1.6 Overview of Systematic Uncertainties

- 2227 Systemic uncertainties are extensively employed to reflect assumptions and inaccura-
2228 cies in the various inputs used by the fit. Two main types of systematic uncertainty
2229 are considered: experimental and modelling. Experimental uncertainties arise due
2230 to the imperfect modelling of the reconstruction algorithms (in particular the jet
2231 reconstruction and b -tagging algorithms), and due to the imperfect modelling of

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%)					
$qg \rightarrow W H \rightarrow \ell\nu b\bar{b}$	POWHEG-Box v2 [138] + GoSam [140] + MiNLO [141, 142]	NNPDF3.0NLO (*) [95]	Pythia 8.212 [97]	AZNLO	NNLO(QCD) + NLO(EW) [143–149]
$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$	POWHEG-Box v2 + GoSAM + MiNLO	NNPDF3.0NLO (*)	Pythia 8.212	AZNLO	NNLO(QCD) (†) + 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 [150–154]
Top quark ($m_t = 172.5$ GeV)					
$t\bar{t}$	POWHEG-Box v2 [138, 155]	NNPDF3.0NLO	Pythia 8.230	A14 [98]	NNLO+NNLL [156]
s -channel	POWHEG-Box v2 [138, 157]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [158]
t -channel	POWHEG-Box v2 [138, 157]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [159]
Wt	POWHEG-Box v2 [138, 160]	NNPDF3.0NLO	Pythia 8.230	A14	Approximate NNLO [161]
Vector boson + jets					
$W \rightarrow \ell\nu$	SHERPA 2.2.1 [162–165]	NNPDF3.0NNLO	SHERPA 2.2.1 [166, 167]	Default	NNLO [168]
$Z/\gamma^* \rightarrow \ell\ell$	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 [130]. If not specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant (α_s). (∗) The events were generated using the first PDF in the NNPDF3.0NLO set and subsequently reweighted to the PDF4LHC15NLO set [169] 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 $qg \rightarrow ZH$ 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 $qg \rightarrow VH$ 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 [170, 171].

pile-up and other effects, as described in Section 7.2. Modelling uncertainties 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 [69].
- The large- R jet energy and mass scales and resolutions. The scales are calibrated as described in Ref. [73], and an uncertainty of 2% and 20% is applied for the jet energy and mass resolutions, respectively.
- b -tagging uncertainties, which are computed separately for b -, c - and light-flavour jets as described in the data-based calibration studies in Refs. [133–135]. 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. [172].
- Uncertainties associated with the reconstruction of the missing transverse energy E_T^{miss} , as described in Ref. [78].
- Uncertainty on the total recorded integrated luminosity, as described in Ref. [40].

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

2259 7.3 Background Modelling Uncertainties

2260 Particular care is paid to the uncertainties on the modelling predictions as discussed
2261 in the following sections. The *Nominal* MC samples are used as a reference to
2262 which different variations can be compared. The nominal samples are chosen as the
2263 best possible representation of the underlying physical process. *Alternative* samples
2264 are used to understand inaccuracies that may be present in the nominal samples.
2265 Some aspect of the nominal model is varied, and the discrepancy with respect to
2266 the nominal model is quantified. The discrepancy is used to estimate a systematic
2267 uncertainty associated with the model parameter which was varied. The alternate
2268 samples are sometimes obtained via internal weight variations or parameterisation
2269 methods, rather than by re-running the simulation. This is discussed in more detail
2270 in Section 7.3.3.

2271 Modelling systematics can have several impacts, including affecting the overall
2272 normalisation for different processes, the relative acceptances between different
2273 analysis regions (i.e. migrations between HP and LP SRs, between the SR and
2274 CR, and between p_T^V bins), and the shapes of the m_J distributions. For each
2275 source of uncertainty, acceptance and shape uncertainties are therefore derived.
2276 Acceptance uncertainties account for the uncertainty in the overall number of events
2277 in each channel, and for the relative numbers of events in the various analysis
2278 regions. Meanwhile, shape uncertainties account for the uncertainty on the kinematic
2279 distributions of the events, but not overall normalisations. They are measured for
2280 the shape of the large- R jet mass distribution, since this is the variable used in the
2281 fit to data.

2282 In this section, truth tagging, the method used to ensure sufficient numbers of jets
2283 are available to calculate uncertainties, is described in Section 7.3.1. The different
2284 sources of background modelling uncertainty which have been assessed are described
2285 in Section 7.3.2, and details of their implementations are found in Section 7.3.3.
2286 Full descriptions of the modelling uncertainties applied for the $V+jets$ and diboson
2287 backgrounds are given in Section 7.3.4 and Section 7.3.5, respectively.

2288 7.3.1 Truth Tagging

2289 Modelling studies involving c - and light-jets is hampered by the low number of events
2290 available after the analysis selection is applied, due to the high rejection rates of the
2291 b -tagging algorithm MV2c10. For modelling studies, truth tagging (TT) is therefore
2292 employed to ensure sufficient numbers of jets are available to calculate uncertainties.
2293 TT works by computing a 2-dimensional efficiency map using the jet p_T and η . The
2294 two leading track-jets associated to the large- R jet are weighted based on their
2295 probability to pass the b -tagging selection, rather than being required to explicitly
2296 pass the b -tagging requirement. The probability to pass the b -tagging selection is
2297 parameterised by the jet p_T and η , and is extracted using a pre-calculated efficiency
2298 map.

2299 7.3.2 Sources of Systematic Modelling Uncertainties

2300 This section briefly describes the different sources of uncertainty in the analysis, and
2301 how each is implemented.

2302 QCD Scales

2303 The $V+jets$ matrix element calculations contains infrared and ultraviolet divergences.
2304 These are handled by introducing arbitrary parameters corresponding to the renormalisation scale (μ_R) and factorisation scale (μ_F). Physical observables are not
2305 dependent on these parameters when using the infinite perturbation series expansion,
2306 however at some fixed order in QCD a limited is present. To assess the impact of
2307 this, both μ_R and μ_F are independently varied from their nominal values by factors
2308 of 0.5 and 2 to account for higher order corrections to the calculation of the matrix
2309 element used to simulate the process.

2311 PDF Sets

2312 Parton distribution functions (PDFs) specify the probability of finding a parton with
2313 a given momentum inside a hadron (in this case, colliding protons). PDFs have
2314 to be derived from data and are a significant source of uncertainty in analyses of

hadronic collision data. There are three sources of 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. [169]. 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 [173] and CT14NLO [174] 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 and 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 (μ_R)	Internal weights
Factorisation scale (μ_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 [175]. 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 $V+hf$ 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
Wbb	$79.3\% \pm 4.4\%$	$75.0\% \pm 8.6\%$	$77.4\% \pm 2.5\%$	$71.7\% \pm 4.5\%$	$68.0\% \pm 7.6\%$	$63.5\% \pm 14.0\%$
Wbc	$6.6\% \pm 0.9\%$	$3.4\% \pm 1.7\%$	$6.2\% \pm 0.5\%$	$5.3\% \pm 0.9\%$	$14.5\% \pm 3.2\%$	$3.4\% \pm 3.2\%$
Wbl	$3.9\% \pm 0.9\%$	$11.4\% \pm 3.5\%$	$4.5\% \pm 0.5\%$	$8.7\% \pm 1.4\%$	$9.8\% \pm 2.2\%$	$9.1\% \pm 3.8\%$
Wcc	$5.1\% \pm 1.7\%$	$6.8\% \pm 2.4\%$	$7.1\% \pm 1.0\%$	$6.3\% \pm 1.4\%$	$4.2\% \pm 2.4\%$	$12.3\% \pm 7.0\%$
Wcl	$2.3\% \pm 1.4\%$	$2.4\% \pm 2.1\%$	$3.4\% \pm 0.7\%$	$5.2\% \pm 1.5\%$	$2.6\% \pm 1.5\%$	$3.4\% \pm 2.1\%$
Wl	$2.9\% \pm 1.0\%$	$0.9\% \pm 1.6\%$	$1.3\% \pm 0.7\%$	$2.8\% \pm 0.7\%$	$0.9\% \pm 0.6\%$	$8.4\% \pm 5.1\%$
Events	187.5 ± 7.7	38.2 ± 3.1	429.5 ± 10.0	97.8 ± 4.2	33.8 ± 2.5	8.3 ± 1.2

Table 7.6: 0-lepton $W+jets$ nominal sample flavour composition and total event yield [176]. 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).

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

Table 7.7: 1-lepton $W+jets$ nominal sample flavour composition and total event yield [176]. 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
Zbb	84.56%	81.84%	82.37%	76.06%	66.12%	63.18%
Zbc	6.03%	6.98%	5.80%	7.46%	15.04%	14.30%
Zbl	4.06%	6.55%	3.83%	6.59%	12.66%	12.81%
Zcc	3.68%	3.40%	5.82%	3.75%	3.36%	3.38%
Zcl	1.23%	0.44%	1.47%	3.97%	1.82%	4.95%
Zl	0.44%	0.78%	0.70%	2.16%	1.00%	1.38%
Events	259.91 ± 4.86	66.12 ± 2.04	420.45 ± 5.73	141.97 ± 2.50	43.49 ± 1.73	16.07 ± 0.83

Table 7.8: 0-lepton $Z+jets$ nominal sample flavour composition and total event yield [176]. 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
Zbb	80.80%	76.95%	79.76%
Zbc	8.10%	6.26%	7.60%
Zbl	4.95%	7.06%	5.52%
Zcc	3.97%	4.46%	4.10%
Zcl	1.61%	3.60%	2.14%
Zll	0.57%	1.68%	0.87%
Events	115.49 ± 2.42	42.42 ± 1.27	157.92 ± 2.73

Table 7.9: 2-lepton $Z+jets$ nominal sample flavour composition and total event yield [176]. 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).

2374 $V + \text{jets}$ Acceptance Uncertainties

2375 Several different types of acceptance uncertainties have been calculated and im-
2376 plemented as nuisance parameters in the fit. These account for the uncertainty
2377 in the overall number of events in each channel, and for the migration of events
2378 between different analysis regions. The acceptance uncertainties relevant to the
2379 $V + \text{jets}$ processes are summarised below.

- 2380 • **Overall normalisation:** only relevant where normalisation cannot be left
2381 unconstrained (or “floating”, i.e. determined as part of the fit). The $V + \text{hf}$
2382 component is left floating in the fit, with independent normalisations used for
2383 $W + \text{hf}$ and $Z + \text{hf}$. The normalisations are mainly determined by the 1-lepton
2384 (for $W + \text{hf}$) and 2-lepton (for $Z + \text{hf}$) regions respectively and then extrapolated
2385 to the 0-lepton channel. The negligible $V + \text{jets}$ backgrounds were constrained to
2386 their cross-sections in the fit.
- 2387 • **SR-to-CR relative acceptance:** the uncertainty on the relative number of
2388 $V + \text{jets}$ events in the signal and control regions.
- 2389 • **HP-to-LP relative acceptance:** the uncertainty on the relative number of
2390 $V + \text{jets}$ events in the HP and LP SRs.
- 2391 • **Medium-to-high p_T^V relative acceptance:** the uncertainty on the relative
2392 number of $V + \text{jets}$ events in the medium and high p_T^V bins.
- 2393 • **Flavour relative acceptance:** for each flavour $V + xx$, where $xx \in \{bc, bl, cc\}$
2394 the ratio of $V + xx/V + bb$ events is calculated. This corresponds to the uncer-
2395 tainty on the heavy flavour composition of the $V + \text{hf}$ background.
- 2396 • **Channel relative acceptance:** the uncertainty on the relative number of
2397 $V + \text{jets}$ events between the channels.

2398 The uncertainties arising from the different sources described in Section 7.3.2 are
2399 summed in quadrature to give a total uncertainty on each region. A summary of
2400 the different acceptance uncertainties that were derived and subsequently applied
2401 in the fit are given in Table 7.10. An effort has been made, wherever possible, to
2402 harmonise similar uncertainties across different analysis regions and channels.

V+jets Acceptance Uncertainties				
Boson	W		Z	
Channel	0L	1L	0L	2L
SR-to-CR	90% [†]	40% [†]	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 † 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 [176].

2403 V +jets Shape Uncertainties

2404 In order to derive shape uncertainties for a given background or signal process,
 2405 normalised distributions of the reconstructed large- R Higgs candidate jet mass m_J
 2406 are compared for the nominal sample and variations. For each variation, the ratio of
 2407 the variation to nominal is calculated, the up and down variations are symmetrised,
 2408 and an analytic function is used to parameterise the ratio. If different analysis regions
 2409 or channels show the same pattern of variation, a common uncertainty is assigned.

2410 An example of a significant source of uncertainty, arising from the choice of factorisa-
 2411 tion scale μ_R is shown in Fig. 7.2. The HP SRs in the medium and high p_T^V bins are
 2412 shown for the 0-lepton channel for the W +hf and Z +hf jet backgrounds. The 0- and
 2413 1-lepton channels for the W +hf contribution and the 0- and 2-lepton channels for
 2414 the Z +jets contribution were found to have compatible shapes in m_J , and so were
 2415 jointly measured. An exponential function $e^{p_0 + p_1 x} + p_2$ has been fitted to the ratio of
 2416 the normalised distributions. The magnitude of the variation is p_T^V dependent, and
 2417 so separate uncertainties are implemented in the fit for each p_T^V region.

2418 The shape uncertainties for μ_R were derived on the SRs but are also applied to the
 2419 CRs, as the low statistics in the CRs make it difficult to derive dedicated shape
 2420 uncertainties. All the shape uncertainties are fully correlated across regions.

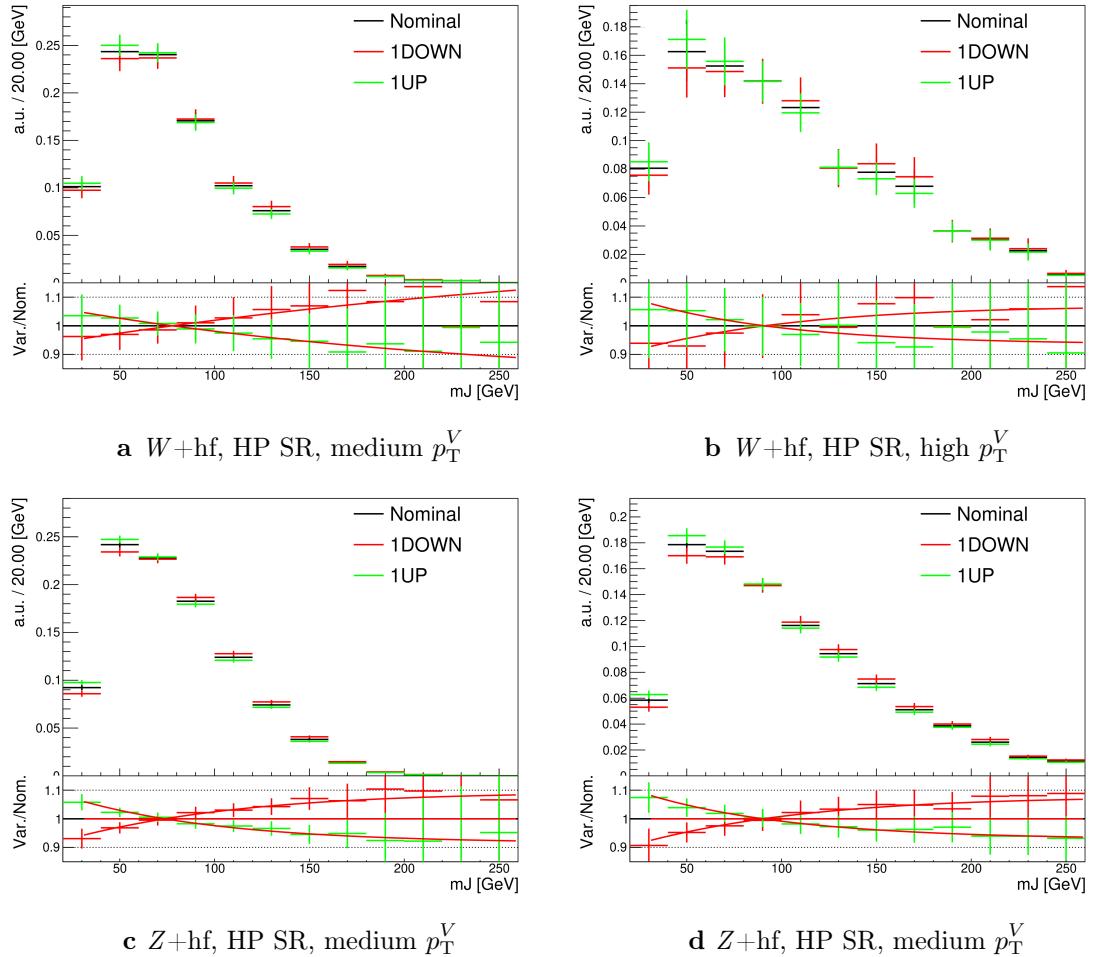


Figure 7.2: Leading large- R jet mass for the Z and $W + hf$ processes in the HP SR of the 0-lepton channel [176]. The renormalisation scale μ_r has been varied by a factor of 0.5 (1DOWN) and 2 (1UP). An exponential function is fitted to the ratio between the nominal and alternate samples.

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 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.

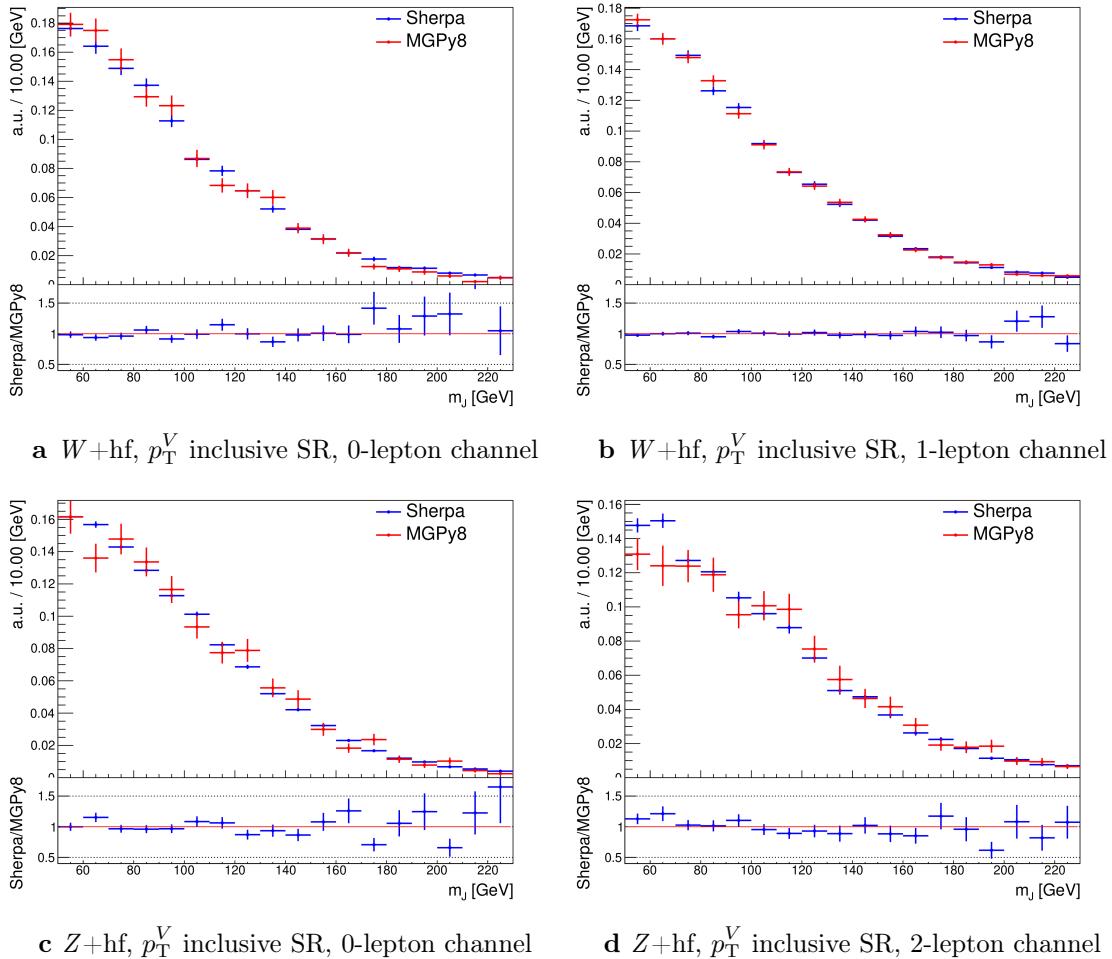


Figure 7.3: Leading large- R jet m_J inclusive in p_T^V for the $V+hf$ process modelled using both the SHERPA +PYTHIA8 (blue) and MADGRAPH +PYTHIA8 (red) samples [176].

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

2429 7.3.5 Diboson Modelling

2430 The procedure to derive the uncertainties for the diboson background generally
2431 follows that of $V+jets$.

2432 The nominal diboson samples are generated using SHERPA 2.2.1 (except for $gg \rightarrow VV$
2433 which uses SHERPA 2.2.2) with the NNPDF3.0NNLO tune. alternate samples were
2434 generated using POWHEG interfaced with PYTHIA8, using the AZNLO shower tune
2435 with the CTEQ6L1 PDFs [177]. Unlike SHERPA, POWHEG models the off-shell Z
2436 contribution at NLO.

2437 Acceptance and shape uncertainties are derived in an analogous fashion to $V+jets$
2438 as described below.

2439 Diboson Acceptance Uncertainties

2440 Diboson acceptance uncertainties are summarised in Table 7.11. Variations from
2441 μ_R , μ_F , PDF choice and an alternative generator are considered and are combined
2442 via a sum in quadrature as described in Section 7.3.4. The largest modification to
2443 the nominal acceptance is from the POWHEG+PYTHIA8 alternate sample. Since
2444 the diboson contribution to the $t\bar{t}$ control region is negligible, no SR-to-CR relative
2445 acceptance uncertainty is necessary.

2446 For the WZ contribution, uncertainties are derived using the 1-lepton channel
2447 and applied to all three channels. The 1-lepton channel was used as it has the
2448 largest amount of available statistics, with the compatibility checked between the
2449 derived uncertainties and the other channels. An additional 8% channel migration
2450 uncertainty is applied on the WZ 0-lepton channel. For the ZZ contribution, the
2451 normalisation uncertainty is calculated using the 2-lepton channel and applied to all
2452 three channels. The 0- and 1-lepton channels were found to have a similar HP-to-LP
2453 relative acceptance uncertainty of 18%. The 1-lepton medium-to-high p_T^V relative
2454 acceptance is estimated using the 2-lepton channel, since the 1-lepton channel had
2455 an insufficient number of events to estimate the uncertainty directly. 30% and
2456 18% channel migration uncertainties are applied in the 0- and 1-lepton channels
2457 respectively.

2458 Since the contribution from WW is negligible, dedicated studies are not performed,
2459 but a 25% normalisation uncertainty is applied in all the three channels which is
2460 based on the modelling studies performed for the previous analysis [79].

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 [176].

2461 Diboson Shape Uncertainties

2462 Diboson shape uncertainties are derived in a similar fashion to $V+jets$. Only the
2463 uncertainties associated with the systematic variations of μ_R and the alternate event
2464 generator have a non-negligible impact on the m_J shape. Variation of μ_R produces
2465 consistent m_J shape changes across all regions and channels, and hence only a single
2466 associated uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted
2467 to the symmetrised ratio.

2468 In the 2-lepton channel, no significant shape differences between the nominal and
2469 alternate was observed. The comparison between the nominal SHERPA and alternate
2470 POWHEG+PYTHIA8 samples is shown in Fig. 7.5 for the 0- and 1-lepton channels
2471 for both the WZ and ZZ processes. For these channels, the shape of m_J varies in
2472 opposite directions in the LP and HP signal regions. Shapes are similar between
2473 p_T^V bins, the 0- and 1-lepton channels and for WZ and ZZ . In order to reduce the
2474 effects of statistical fluctuations on the fitted shape, these regions are merged before
2475 deriving the shape uncertainty. A third order polynomial is fitted to the ratio, and
2476 this function transitions to a constant piecewise function in the high mass region
2477 to accurately represent the shape. Dependence on the event generator was found
2478 to be negligible within statistical uncertainty in the 2-lepton channel, and so no

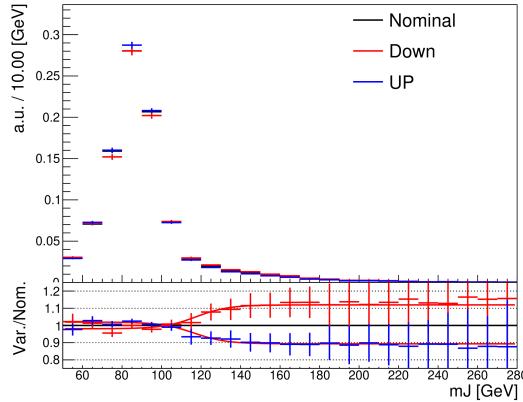


Figure 7.4: Leading large- R jet mass distribution for the combined WZ and ZZ processes, inclusive across all signal regions and lepton channels [176]. The renormalisation scale μ_R has been varied by a factor of 2 (UP) and 0.5 (DOWN). The red and blue curves show the fitted results of the hyperbolic tangent function.

2479 uncertainty was applied. All diboson shape uncertainties are fully correlated in the
2480 fit.

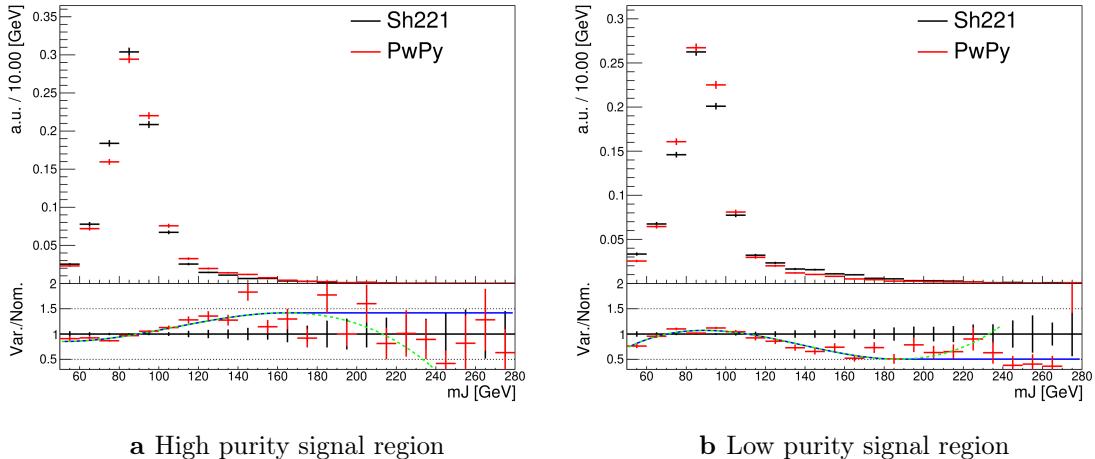


Figure 7.5: Comparison of the shapes of the large- R jet mass m_J between SHERPA (Sh221) (black) and POWHEG+PYTHIA8 (PwPy) (red) samples from the WZ and ZZ processes in high and low purity signal regions, integrated over p_T^V regions and the 0- and 1-lepton channels [176]. The dashed green line shows 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. [176].

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. [178] 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.

2504 The largest sources of acceptance and shape uncertainties were due to this DS-DR
2505 variation. Acceptance uncertainties for the single-top background are enumerated in
2506 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. [176].

2507 7.3.7 Signal Modelling

2508 The modelling of the systematic uncertainties affecting the signal processes follows
2509 the procedure described in Refs. [26, 179, 180]. The $qq \rightarrow VH$ signal samples are
2510 generated with POWHEG BOX v2 +GoSAM at next-to-leading order (NLO) accuracy
2511 in QCD. An additional $gg \rightarrow ZH$ sample is generated using POWHEG BOX v2 at
2512 leading order (LO) in QCD. In both cases, the generated events are interfaced with
2513 PYTHIA 8 for the parton showering modelling. An alternate HERWIG7 sample is used
2514 to assess the uncertainty on the parton showering model. Recommended systematic
2515 uncertainties on the signal production cross-sections and $H \rightarrow b\bar{b}$ branching ratio
2516 from the LHC Higgs Cross Section Working Group are applied [181, 182]. This
2517 includes acceptance and shape uncertainties arising from missing higher-order QCD
2518 and electroweak corrections, PDF uncertainties, renormalisation and factorisation
2519 scales, and an alternate parton showering model.

2520 7.4 Statistical Treatment

2521 A binned global maximum-profile-likelihood fit of the m_J distribution is performed
 2522 to extract information on the signal, combining all the analysis regions defined
 2523 in Table 7.1. The signal strength $\mu = \sigma/\sigma_{\text{SM}}$ is defined as the ratio between the
 2524 observed and predicted cross-sections, where $\mu = 0$ corresponds to the background-
 2525 only hypothesis and $\mu = 1$ corresponds to the SM prediction. This is a parameter of
 2526 interest (POI) which acts to scale the total number of signal events, and is determined
 2527 during the fit procedure.

2528 The present analysis makes use of two POIs. The first, μ_{VH}^{bb} , is the signal strength
 2529 for the VH , $H \rightarrow b\bar{b}$ process, the primary process under investigation. The diboson
 2530 production strength μ_{VZ}^{bb} for the VZ , $Z \rightarrow b\bar{b}$ process is measured simultaneously and
 2531 provides a validation of the analysis apparatus used for the primary $H \rightarrow b\bar{b}$ measure-
 2532 ment. Alongside the two POIs, the predictive model depends on several parameters
 2533 which are not the primary target of measurement, and represent the systematic
 2534 uncertainties discussed previously. These parameters are called nuisance parameters
 2535 (NPs), collectively referred to as θ . Freely floating background normalisations are
 2536 implemented as NPs and are also extracted during the fitting processes.

2537 7.4.1 The Likelihood Function

2538 The statistical setup treats each bin as a Poisson counting experiment and is based on
 2539 the ROOSTATS framework [183]. The combined likelihood over N bins is constructed
 2540 as the product of Poisson probabilities in each bin. Considering the simplified case
 2541 of a single signal strength parameter μ , and neglecting sources of systematic or
 2542 statistical uncertainty, this is given by

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

2543 where s_i (b_i) is the expected number of signal (background) events in bin i , and n_i is
 2544 the number of observed data events in bin i .

2545 **Treatment of Uncertainties**

2546 Systematic uncertainties can modify the predicted signal and background yields s_i
2547 and b_i . Each source of systematic uncertainty is taken into account by adding an
2548 additional NP θ_j to the likelihood in the form of a Gaussian cost function. The
2549 combined effect of the NPs is then

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

2550 where N_θ is the number of NPs, θ_j is the nominal value of the j th NP, $\hat{\theta}_j$ is the fitted
2551 value, and σ_j is the corresponding associated prior uncertainty on the NP. As the
2552 fitted value $\hat{\theta}_j$ deviates from its nominal value, a cost is introduced. The presence of
2553 NPs modifies the likelihood as

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

2554 The predicted signal and background yields are also modified by the presence of the
2555 NPs with

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

2556 For NPs which are left freely floating in the fit, no corresponding Gaussian constraints
2557 are added to the likelihood.

2558 The pull of a NP is defined as the difference between the fitted value $\hat{\theta}_j$ and the
2559 nominal value θ_j , divided by the uncertainty on the NP σ_j . To obtain the uncertainty
2560 on the pull of a NP, the following procedure is used. The Hessian matrix \mathbf{H} is
2561 calculated as

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

2562 Taking the inverse of the Hessian matrix \mathbf{H}^{-1} yields the covariance matrix, from
2563 which the post-fit uncertainties on the different NPs can be extracted. If the post-fit
2564 uncertainty is smaller than the nominal uncertainty, additional information about
2565 the NP has been extracted by the fit, and NP is said to be *constrained*.

2566 Following Ref. [184], the statistical uncertainty on the simulated events is implemented
2567 using a dedicated NP for each bin which can scale the background yield in that bin.
2568 Statistical NPs are also implemented using a Gaussian constraint.

2569 Smoothing and Pruning

2570 To simplify the fit to reduce and improve its robustness, systematic uncertainties
2571 are smoothed and pruned. Smoothing accounts for the large statistical uncertainty
2572 present in some samples that can lead to unphysical fluctuations in the shape
2573 systematics. The smoothing procedure relies on the assumption that the impact of
2574 systematics should be approximately monotonic and correlated between neighbouring
2575 bins.

2576 In addition to smoothing, pruning is the process of removing from the fit those
2577 systematics which only have a very small effect. This improves the stability of the
2578 fit by reducing the number of degrees of freedom. Acceptance uncertainties are
2579 pruned in a given region if they have a variation of less than 0.5%, or if the up and
2580 down variations have the same sign in that region. Shape uncertainties are pruned
2581 in a given region if the deviation in each bin is less than 0.5% in that region. In
2582 addition, acceptance and shape uncertainties are neglected in a given region for any
2583 background which makes up less than 2% of the total background in that region.

2584 Fit Procedure and Statistical Tests

2585 The best-fit value of μ , denoted $\hat{\mu}$, is obtained via an unconditional maximisation of
 2586 the likelihood. The likelihood is also used to construct a statistical test which can
 2587 confirm or reject the background-only hypothesis. The test statistic q_μ is constructed
 2588 from the profile likelihood ratio,

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

2589 where $\hat{\mu}$ and $\hat{\theta}$ are chosen to maximise the likelihood \mathcal{L} , and the profile value $\hat{\theta}_\mu$ is
 2590 obtained from a conditional maximisation of the likelihood for a specific choice of
 2591 $\mu = 0$ corresponding to the background-only hypothesis.

2592 The test statistic is used to construct a p -value which is used to probe the background-
 2593 only hypothesis. The p -value is typically reported in terms of the significance Z ,
 2594 defined as the number of standard deviations for a Gaussian Normal distribution
 2595 which will produce a one-sided tail integral equal to the p -value, as in

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

2596 Typically a value of $Z = 3$ constitutes *evidence* of a processes, while $Z = 5$ is
 2597 required for a *discovery*, or *observation*. Alongside the p -value, the best-fit value of
 2598 the signal strength $\hat{\mu}$ and its corresponding uncertainty are quoted, and compared to
 2599 their expected values. More detail on the statistical methodology can be found in
 2600 Ref. [185].

2601 7.4.2 Background Normalisations

2602 The backgrounds which can be constrained by the fit are left freely floating and
 2603 the corresponding normalisation factors are extracted. Normalisation factors (NF),
 2604 represent the value by which the predicted normalisations are scaled, and are im-
 2605 plemented for the dominant backgrounds ($t\bar{t}$, $Z+hf$, $W+hf$). The NFs are also
 2606 subdivided into different regions of phase-space for $t\bar{t}$, given it is possible to obtain

2607 a strong constraint in the individual channels. This also removes the need for an
2608 extrapolation uncertainty.

2609 The normalisations and shapes of all other backgrounds, with the exception of the
2610 multijet background which is estimated using a data driven technique, are initialised
2611 using the nominal samples and the state-of-the art process normalisations, as outlined
2612 in Table 7.4.

2613 7.4.3 Asimov Dataset & Expected Results

2614 The Asimov dataset is constructed by replacing the data with the sum of the signal
2615 and background predictions $n_i = s_i + b_i$. A fit to this dataset using the nominal
2616 values of the NPs from the simulation will recover the input values and is useful for
2617 studying the expected result, in addition to constraints on and correlations between
2618 the NPs.

2619 Alternatively, a conditional fit to the Asimov dataset can be performed using values
2620 of the background NPs which are determined from an unconditional fit to data. The
2621 signal NPs and POIs are fixed at their nominal values from the SM simulation. The
2622 result of this fit can be used to calculate expected (median) significances given a
2623 more realistic background model, which can be compared to their observed values,
2624 as is done in Section 7.5.2.

2625 7.5 Results

2626 In the present analysis, the two signal strength parameters μ_{VH}^{bb} and μ_{VZ}^{bb} are extracted
2627 from a simultaneous maximisation of the likelihood described in Section 7.4. The
2628 results of the analysis are summarised in this section. Post-fit m_J distributions are
2629 shown in Section 7.5.1. The observed signal strengths are given in Section 7.5.2,
2630 along with observed and expected significances. Finally in Section 7.5.3 the impact
2631 of systematic uncertainties on the results is examined.

2632 7.5.1 Post-fit Results

2633 In addition to the observed significance and signal strength, it is also necessary to
 2634 study the post-fit m_J yields and distributions to compare the level of the agreement
 2635 between the simulation (using the best-fit values of the signal strength $\hat{\mu}$ and the
 2636 NP $\hat{\theta}$) and the data. The best-fit values $\hat{\mu}$ and $\hat{\theta}$ are obtained from an unconditional
 2637 fit to data over all analysis regions. The post-fit background normalisation factors
 2638 extracted from the unconditional fit to data fit are shown in Table 7.14, and the
 2639 post-fit yields are presented in Table 7.15, Table 7.16, and Table 7.17 for the 0-, 1-
 2640 and 2-lepton channels, respectively.

Process	Normalisation factor
$t\bar{t}$ 0-lepton	0.88 ± 0.10
$t\bar{t}$ 1-lepton	0.83 ± 0.09
$W+hf$	1.12 ± 0.14
$Z+hf$	1.32 ± 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 [130]. 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 ± 11.17	18.99 ± 9.76	1.05 ± 0.54	5.69 ± 2.88	5.85 ± 3.01	0.33 ± 0.17
$W+t$	14.70 ± 5.37	45.55 ± 19.44	17.18 ± 8.09	2.03 ± 0.98	8.93 ± 6.33	3.76 ± 2.49
other $t+X$	0.79 ± 0.03	3.18 ± 0.66	4.51 ± 1.28	-	0.66 ± 0.03	0.11 ± 0.00
$t\bar{t}$	75.19 ± 13.60	423.85 ± 36.12	539.21 ± 31.39	7.54 ± 1.77	38.20 ± 6.75	44.07 ± 7.43
VZ	77.01 ± 17.09	87.70 ± 19.36	6.16 ± 1.56	17.30 ± 4.10	28.77 ± 6.55	2.79 ± 0.72
WW	-	2.15 ± 0.05	0.24 ± 0.01	0.33 ± 0.02	1.80 ± 0.06	-
Whf	100.78 ± 20.01	331.31 ± 59.54	29.97 ± 21.85	20.19 ± 6.24	59.82 ± 17.91	6.61 ± 5.09
Wcl	5.13 ± 2.31	8.44 ± 3.24	0.46 ± 0.01	0.99 ± 0.69	2.77 ± 1.14	0.19 ± 0.07
Wl	5.61 ± 3.93	4.61 ± 2.45	0.16 ± 0.00	1.41 ± 2.06	2.67 ± 1.67	0.57 ± 0.36
Zhf	318.76 ± 35.27	548.71 ± 61.84	76.97 ± 21.47	86.79 ± 10.63	184.99 ± 21.43	25.76 ± 7.43
Zcl	3.97 ± 1.63	6.74 ± 2.68	0.83 ± 0.02	-	6.36 ± 2.73	0.93 ± 0.41
Zl	1.34 ± 0.67	3.61 ± 2.14	0.42 ± 0.01	1.05 ± 0.63	3.68 ± 2.47	0.29 ± 0.16
Data	623	1493	683	146	330	85
Background	603 ± 25	1466 ± 36	676 ± 25	138 ± 9	339 ± 15	85 ± 7

Table 7.15: Post-fit yields in the 0-lepton channel. Combined statistical and systematic uncertainties are shown [176].

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

Table 7.16: Post-fit yields in the 1-lepton channel. Combined statistical and systematic uncertainties are shown [176].

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

Table 7.17: Post-fit yields in the 2-lepton channel. Combined statistical and systematic uncertainties are shown [176].

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.

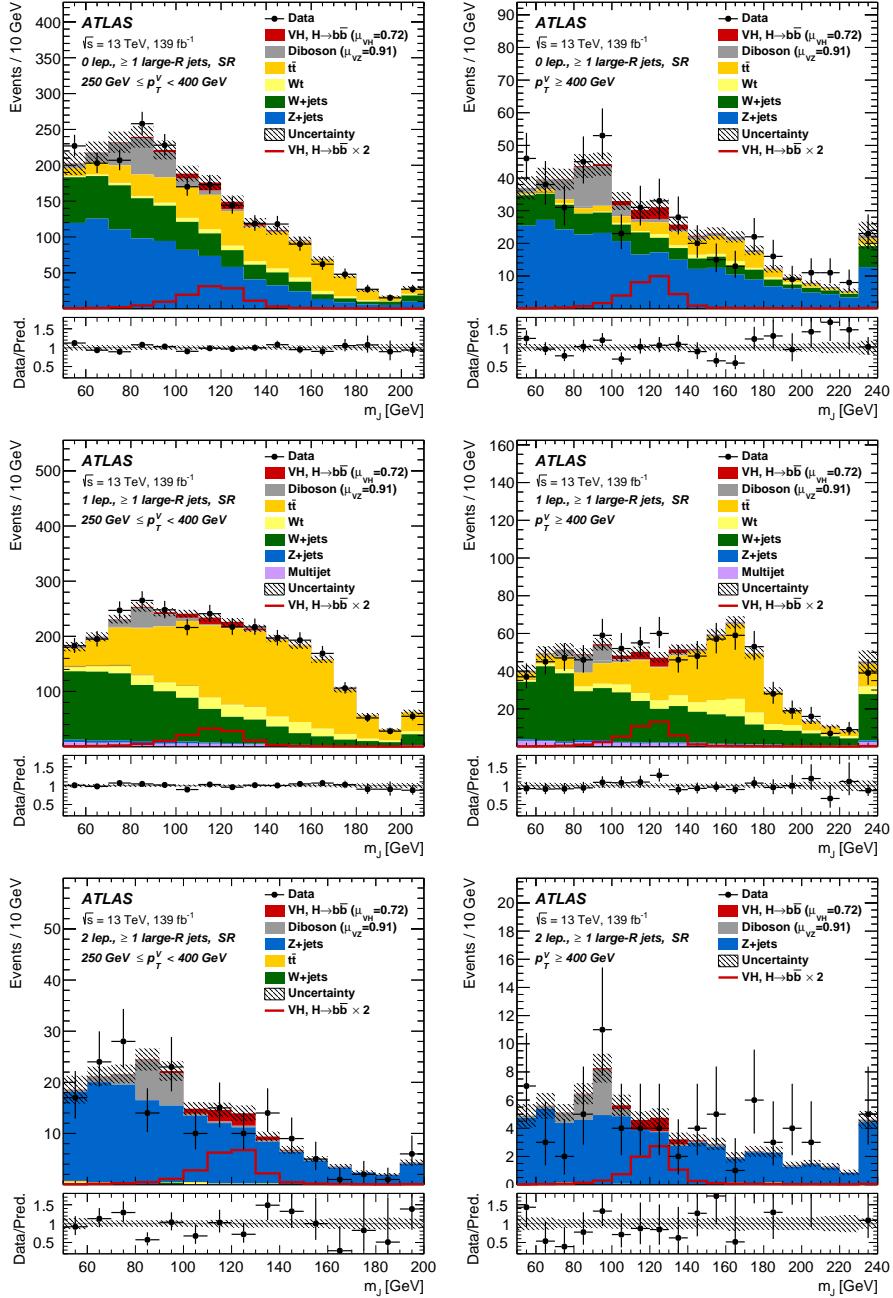


Figure 7.6: The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$. The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions 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 [130].

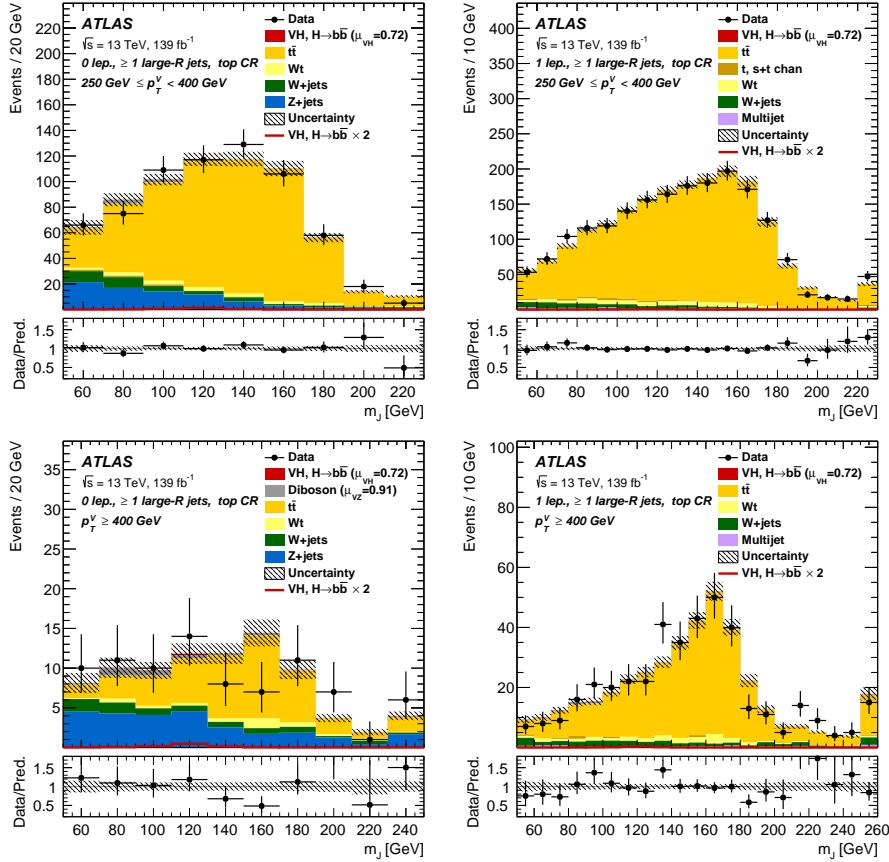


Figure 7.7: The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. The background contributions after 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 [130].

2647 7.5.2 Observed Signal Strength & Significance

2648 The measured signal strength is computed as the ratio between the measured signal
 2649 yield to the prediction from the SM. The combed result for all three lepton channels
 2650 and all analysis regions is given for μ_{VH}^{bb} in Eq. (7.8), and for μ_{VZ}^{bb} in Eq. (7.9). Both
 2651 results include a full breakdown of the systematic and statistical uncertainties.

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

2652

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

2653 The results for μ_{VH}^{bb} and μ_{VZ}^{bb} are consistent with the expectation from the SM. The
 2654 μ_{VH}^{bb} measurement is dominated by statistical uncertainty, while the μ_{VZ}^{bb} measurement
 2655 is dominated by systematic sources of uncertainty. The measured signal strength
 2656 for μ_{VZ}^{bb} corresponds to an observed significance of 2.1 standard deviations, with an
 2657 expected (median) significance given the SM prediction of 2.7 standard deviations.
 2658 The diboson observed (expected) signal strength significance is 5.4 (5.7). These results
 2659 are summarised in Fig. 7.8, which shows the weighted and background-subtracted m_J
 2660 distribution. A clear signal excess is visible around the Higgs mass of $m_H = 125$ GeV.

2661 Compatibility Studies

2662 Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting μ_{VH}^{bb}
 2663 into three separate POIs, one for each channel. A simultaneous fit to the channel
 2664 specific signal strengths can then be performed, which allows a comparison of the
 2665 contributions from each channel. Fig. 7.9 compares the best-fit signal strengths.
 2666 The 0- and 1-lepton channels show a signal strength which is consistent with the
 2667 SM prediction, while the 2-lepton channel shows a small deviation within the 1σ
 2668 uncertainty. Overall, good compatibility is observed via a χ^2 test with a corresponding
 2669 p -value of 49%.

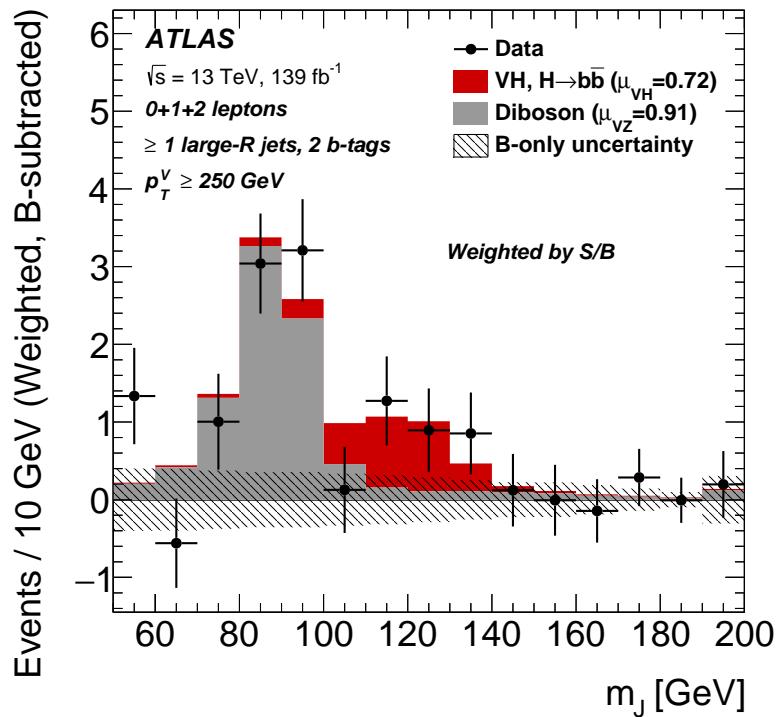


Figure 7.8: m_J distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected 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 [130].

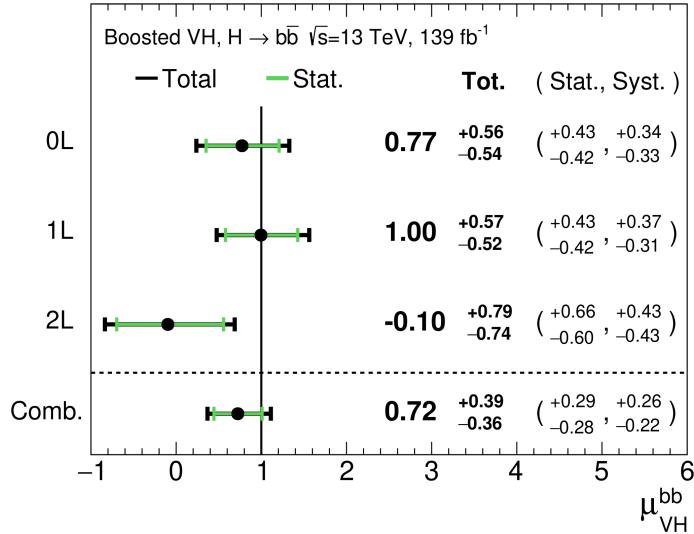


Figure 7.9: Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels fit separately) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a χ^2 difference test and results in a p-value of 49% [130].

2670 7.5.3 Impact of Systematics

2671 The impact of systematic uncertainties on the final fitted value $\hat{\mu}^{bb}$ can be studied
2672 using the NP rankings, and the uncertainty breakdown.

2673 Fig. 7.10 shows the NP ranking, which is used to visualise which NPs have the largest
2674 impact on the sensitivity to the fitted POI. To obtain the ranking, a likelihood scan
2675 is performed for each NP θ_j . First, an unconditional fit is used to determine $\hat{\theta}_j$.
2676 The NP is then fixed to its post-fit value varied by $\pm 1\sigma$, the fit is repeated and the
2677 difference between the best-fit value of the POI, $\Delta\hat{\mu}_{VH}^{bb}$, is calculated, and used to
2678 rank the NPs. In addition, the pulls and constraints for the highest ranked NPs are
2679 also shown.

2680 The experimental uncertainty on the signal large- R jet mass resolution (JMR) has
2681 the largest impact. JMR and jet energy scale (JES) uncertainties also have large
2682 impacts for the $V+jets$ and for the diboson backgrounds. The freely-floating $Z+hf$
2683 normalisation is the second highest ranked NP, and is heavily constrained by the
2684 fit. The VZ POI μ_{VZ}^{bb} is also a significant NP when considering the primary μ_{VH}^{bb}
2685 measurement.

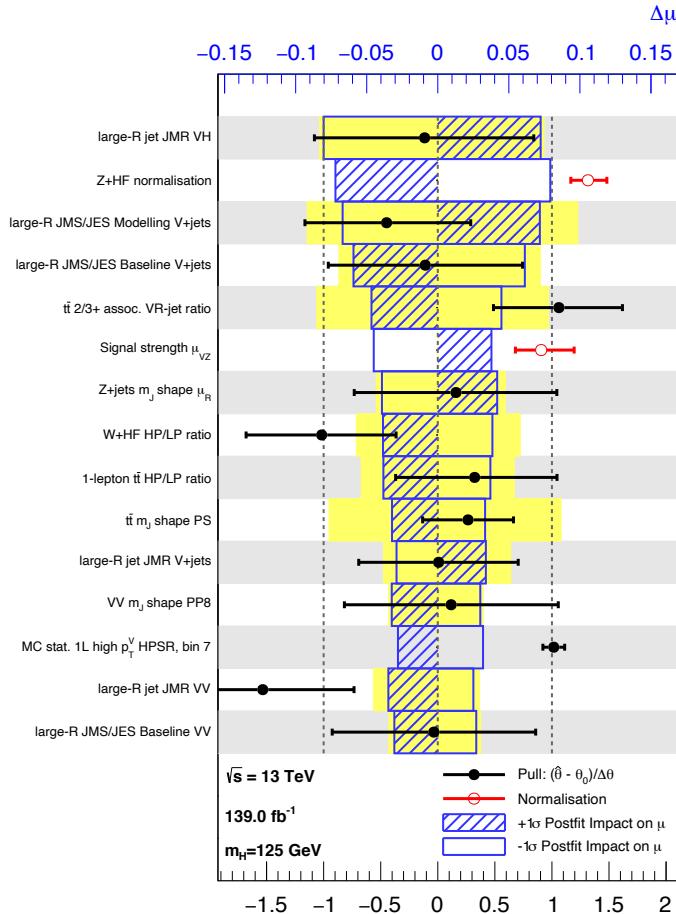


Figure 7.10: Impact of systematic uncertainties on the fitted VH signal-strength parameter $\hat{\mu}_{VH}^{\text{bb}}$ sorted in decreasing order. The boxes show the variations of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual nuisance parameter to its post-fit value 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 ± 1 . The filled circles show the deviation of the fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value θ_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 [176].

2686 The NP ranking highlights individual NPs which have a large impact on the POI
 2687 measurement sensitivity. Complementary information is provided at a higher level
 2688 by considering the overall impact of different groups of systematics. The groups
 2689 are constructed from NPs which have similar physical origins. The impact of each
 2690 group is calculated by running a fit with all the NPs in the given group fixed to their
 2691 nominal values. The uncertainty on the POI extracted from this fit is subtracted in
 2692 quadrature from the uncertainty on the POI from the nominal fit, and the resulting
 2693 values are provided as the impact for each group. The total systematic impact
 2694 is the difference in quadrature between the nominal uncertainty on μ_{VH}^{bb} and the
 2695 estimated impact for the combined statistical uncertainties. The “data stat only”
 2696 group fixes all NPs at their nominal value, while the total statistical impact fixes all
 2697 NPs except floating normalisations. The floating normalisations group fixes only the
 2698 NPs associated with normalisation which are left floating in the fit.

2699 The full breakdown for the observed impact of uncertainties on the μ_{VH}^{bb} signal
 2700 strength is provided in Table 7.18. The uncertainty on μ_{VH}^{bb} is dominated by com-
 2701 bined statistical effects (0.28), although the combined impact of systematics (0.24) is
 2702 of a comparable size. The signal largest group is the data stat uncertainty (0.25),
 2703 demonstrating that the analysis would benefit from an increased integrated lumi-
 2704 nosity or improved efficiency to select signal events (recall from Section 7.1.2 the
 2705 signal efficiency is in the range of 10%). Of the experimental systematic sources of
 2706 uncertainty, the dominant impact is from the experimental uncertainties associated
 2707 with the simulation of large- R jets (0.13). Other experimental sources of uncertainty
 2708 are small in comparison. Modelling uncertainties also have a large contribution to the
 2709 overall systematic uncertainty. The biggest contribution to the overall uncertainty
 2710 is the combined statistical uncertainty on the simulated samples (0.09). Out of the
 2711 backgrounds, the $W+jets$ and $Z+jets$ have the highest (0.06) and second-highest
 2712 (0.05) impact respectively.

2713 7.5.4 STXS Interpretation

2714 The Simplified Template Cross Sections (STXS) framework provides a common
 2715 categorisation of candidate Higgs boson events according to certain truth-level
 2716 properties of the production mode under study [26, 186]. 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^{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 μ_{VH}^{bb} obtained from the (1+1)-POI fit. The average impact represents the average between the positive and negative uncertainties on μ_{VH}^{bb} . The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations [176].

2717 designed to be independent of the decay mode of the Higgs boson, and is therefore
2718 well suited to the combination of measurements between different decay channels
2719 and experiments.

2720 The STXS cross sections are independently measured for the ZH and WH production
2721 modes following the approach described in [179]. For each production mode, two
2722 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.
2724 Events from the simulated signal samples are categorised into the regions and used
2725 to estimate the expected cross section times branching ratio $\sigma \times B$ in each region,
2726 where

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

2727 A simultaneous fit of the four cross section times branching ratios is performed.
2728 The uncertainties described in Section 7.3 are reused for the STXS fit, with the
2729 exception of the theoretical uncertainties on the signal cross section and branching
2730 ratios. The result from the fit is shown in Section 7.5.4 and compared with the
2731 expected prediction from the SM. The expected and observed results agree within
2732 the given uncertainties.

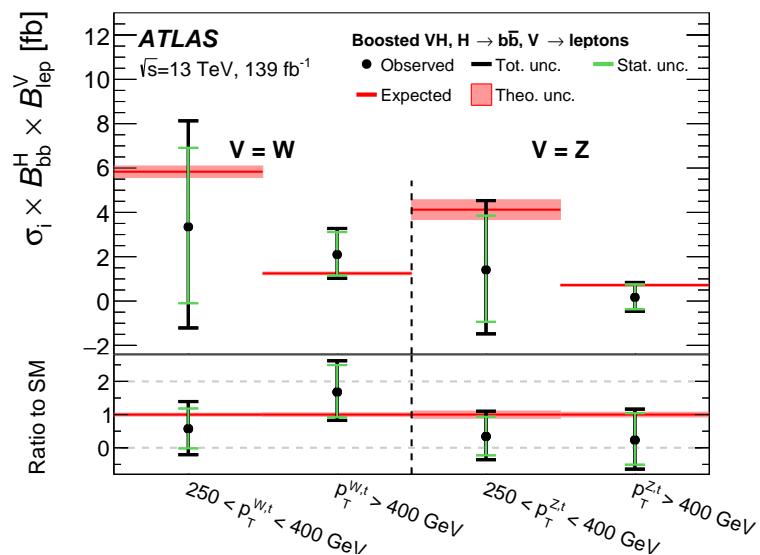


Figure 7.11: Measured VH simplified template cross sections times the $H \rightarrow b\bar{b}$ and $V \rightarrow \text{leptons}$ branching fractions in the medium and high $p_T^{V,t}$ bins [130].

2733 7.6 Conclusion

2734 The analysis of the associated production of vector bosons with boosted Higgs bosons
2735 decaying to a pair of b -quarks using large- R jets is presented. The Higgs candidate is
2736 reconstructed as a large- R jet in order to improve sensitivity in the boosted regime
2737 in which the Higgs decay products are significantly collimated. The analysis is
2738 performed using 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13\text{ TeV}$ collected
2739 throughout the duration of Run 2 of the LHC.

2740 In comparison with the null hypothesis, the Standard Model (SM) VH , $H \rightarrow b\bar{b}$
2741 process is found to have an observed significance of 2.1 standard deviations, whereas
2742 the corresponding expected significance is 2.7 standard deviations. The VH , $H \rightarrow b\bar{b}$
2743 process is measured simultaneously with the diboson VZ , $Z \rightarrow b\bar{b}$ process, which
2744 provide a cross-check for the main analysis. The observed (expected) significance for
2745 the diboson process is 5.4 (5.7).

2746 The results are interpreted in the context of the STXS framework. The cross sections
2747 for the WH and ZH processes are measured in two p_T^V bins, and are found to agree
2748 with the SM prediction within the given uncertainties. At the time of publication,
2749 the results are the most precise measurements of the WH and ZH , $H \rightarrow b\bar{b}$ cross
2750 sections in the high p_T^V regime.

2751 The statistical and systematic sources of uncertainty contribute a similar amount
2752 to the overall uncertainty on the result. This analysis would therefore likely benefit
2753 greatly from the improved b -tagging efficiency at high- p_T enabled by GN1 as discussed
2754 in Chapter 6, due to the associated reduction in statistical uncertainty provided by
2755 the increased number of events used in the analysis.

2756 The large- R jet mass resolution is found to be the dominant source of systematic
2757 uncertainty on the μ_{VH}^{bb} measurement. An improved method of reconstructing the
2758 large- R jet mass, for example by using a machine learning based regression approach,
2759 possibly as an additional auxiliary task to GN1 (see Chapter 6), could reduce the
2760 systematic uncertainty on the μ_{VH}^{bb} measurement. Statistical uncertainty could be
2761 reduced by increasing the integrated luminosity used to perform the analysis by
2762 combining with Run 3 data.

2763 **Chapter 8**

2764 **Conclusion**

2765 **8.1 Summary**

2766 The current understanding of particle physics contains many unanswered questions,
2767 and improving our understanding of the Standard Model is a promising way to
2768 attempt to answer some of them. One of the key particles which may enhance this
2769 understanding is the Higgs boson, which was first observed only a decade ago and
2770 remains under intense scrutiny at the LHC. Given it's propensity to decay to heavy
2771 flavour b -quarks, reconstructing and identifying b -jets is of crucial importance to
2772 improving our understanding in this area. As discussed in Chapter 4, this task
2773 becomes increasingly difficult at high transverse momenta.

2774 One of the effects that hampered tracking and b -tagging performance at high- p_T was
2775 identified to be the increased rate of fake tracks. To address this issue, an algorithm
2776 was developed which is able to successfully identify fake tracks within jets 45% of
2777 the time, with a minimal loss of signal tracks of 1.2%. Removal of such tracks was
2778 found to improve the light-jet mistagging rate of the SV1 and JetFitter algorithms
2779 by up to 20% at high transverse momentum.

2780 A novel approach to b -tagging, GN1 was also developed using a Graph Neural Network
2781 (GNN) architecture. The model is encouraged to learn the topology of the jet through
2782 vertexing and track classification auxiliary tasks. As a single end-to-end trained
2783 model, GN1 simplifies the complexity of the flavour tagging pipeline and is able to
2784 achieve superior performance to the current state-of-the-art algorithms, which rely
2785 on a two-tiered approach. Compared with DL1r, GN1 improves the light-jet rejection

2786 by a factor of ~ 1.8 for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ at the 70%
2787 b -jet WP and by a factor of ~ 6 for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$
2788 for a corresponding b -jet efficiency of 30%. GN1 also demonstrates a significant
2789 improvement in the discrimination between b - and c -jets, which also contributes to
2790 improved c -tagging performance. In the high- p_T regime, GN1 improves the b -jet
2791 tagging efficiency by 100% for a fixed light-jet rejection of 100. Initial validation of
2792 the model in data has been performed. The level of agreement between the simulation
2793 and data that is observed is similar to previous flavour tagging algorithms. GN1 has
2794 been successfully deployed in the ATLAS High Level Trigger, and shows promising
2795 performance when trained on high pile-up samples corresponding to the HL-LHC
2796 conditions. Ultimately the improved jet tagging performance enabled by the new
2797 algorithm will have a large impact across a broad spectrum of the ATLAS physics
2798 programme.

2799 This thesis demonstrates that even with suboptimal track reconstruction in this
2800 regime, it is possible to make algorithmic advancements to the flavour tagging pipeline
2801 to improve the identification of b -jets. This work has impacts for any analysis which
2802 relies on the identification of b -jets, including those which are sensitive to the Higgs
2803 boson.

2804 Analysis of VH , $H \rightarrow b\bar{b}$ events was also carried out with 139 fb^{-1} of Run 2
2805 ATLAS at $\sqrt{s} = 13 \text{ TeV}$. Various background modelling uncertainties were derived
2806 and investigations into the fit model were carried out. The analysis observed a
2807 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
2808 observed (expected) significance of 2.1σ (2.7σ). The result was validated using a
2809 simultaneous fit to the VZ , $Z \rightarrow b\bar{b}$ process, which acts as a cross check to validate
2810 the primary analysis. The results of the analysis are the most precise measurements
2811 available in the high- p_T for the VH , $H \rightarrow b\bar{b}$ process. The high- p_T region is of
2812 particular interest as it is a region of phase space with good sensitivity to new physics.

2813 8.2 Future Work

2814 Additional algorithmic improvements are likely to yield further improved flavour
2815 tagging performance. Aside from these, large improvements to the flavour tagging
2816 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_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 6.6. 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 MVA 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.

2850 **Bibliography**

- 2851 [1] A. Buckley, *A class for typesetting academic theses*, (2010). [http://ctan.tug.org/
2852 tex-archive/macros/latex/contrib/heptesis/heptesis.pdf](http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf).
- 2853 [2] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,
2854 *Secondary vertex finding in jets with neural networks*, *The European Physical Journal C* 81 (2021) .
2855
- 2856 [3] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS
2857 Detector*, ATL-PHYS-PUB-2022-027 (2022).
- 2858 [4] L. Morel, Z. Yao, P. Cladé and S. Guellati-Khélifa, *Determination of the
2859 fine-structure constant with an accuracy of 81 parts per trillion*, *Nature* 588 (2020)
2860 pp. 61–65.
- 2861 [5] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al.,
2862 *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* 606
2863 (2022) pp. 479–483.
- 2864 [6] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev.
2865 Lett.* 74 (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- 2866 [7] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* 74 (1995)
2867 pp. 2632–2637 [[hep-ex/9503003](#)].
- 2868 [8] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV
2869 Proton-Nucleus Collisions*, *Phys. Rev. Lett.* 39 (1977) pp. 252–255.
- 2870 [9] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy
2871 Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* 122
2872 (1983) pp. 103–116.
- 2873 [10] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* 504
2874 (2001) pp. 218–224 [[hep-ex/0012035](#)].

- 2875 [11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
2876 *Phys. Rev. Lett.* 13 (1964) pp. 321–323.
- 2877 [12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.*
2878 13 (1964) pp. 508–509.
- 2879 [13] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*
2880 *Massless Particles*, *Phys. Rev. Lett.* 13 (1964) pp. 585–587.
- 2881 [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
2882 *Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* 716 (2012)
2883 p. 1 [[1207.7214](#)].
- 2884 [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*
2885 *experiment at the LHC*, *Phys. Lett. B* 716 (2012) p. 30 [[1207.7235](#)].
- 2886 [16] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* 2022 (2022)
2887 p. 083C01.
- 2888 [17] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge*
2889 *invariance*, *Phys. Rev.* 96 (1954) pp. 191–195.
- 2890 [18] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* 22 (1961)
2891 pp. 579–588.
- 2892 [19] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* 19 (1967) pp. 1264–1266.
- 2893 [20] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel*
2894 *symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, Conf. Proc. C680519
2895 (1968) pp. 367–377.
- 2896 [21] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,
2897 *Phys. Rev.* 104 (1956) pp. 254–258.
- 2898 [22] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hopper and R. P. Hudson, *Experimental*
2899 *test of parity conservation in beta decay*, *Phys. Rev.* 105 (1957) pp. 1413–1415.
- 2900 [23] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of*
2901 *conservation of parity and charge conjugation in meson decays: the magnetic moment*
2902 *of the free muon*, *Phys. Rev.* 105 (1957) pp. 1415–1417.
- 2903 [24] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,
2904 *Phys. Rev.* 117 (1960) pp. 648–663.

- 2905 [25] J. Goldstone, *Field theories with « superconductor » solutions*, *Il Nuovo Cimento*
2906 (1955-1965) 19 (1961) pp. 154–164.
- 2907 [26] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs*
2908 *Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [[1610.07922](#)].
- 2909 [27] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2910 *ATLAS detector*, [ATLAS-CONF-2018-036](#) (2018).
- 2911 [28] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys.*
2912 *Rev. Lett.* 121 (2018) p. 121801 [[1808.08242](#)].
- 2913 [29] L. Evans and P. Bryant, *LHC Machine*, *JINST* 3 (2008) p. S08001.
- 2914 [30] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of*
2915 *Instrumentation* 3 (2008) pp. S08002–S08002.
- 2916 [31] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* 3 (2008)
2917 p. S08004.
- 2918 [32] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of*
2919 *Instrumentation* 3 (2008) pp. S08005–S08005.
- 2920 [33] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
2921 *JINST* 3 (2008) p. S08003.
- 2922 [34] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data*
2923 *taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 2925 [35] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*,
2926 (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 2928 [36] CERN, *The accelerator complex*, (2012).
2929 <http://public.web.cern.ch/public/en/research/AccelComplex-en.html>.
- 2930 [37] G. C. Strong, *On the impact of selected modern deep-learning techniques to the*
2931 *performance and celerity of classification models in an experimental high-energy*
2932 *physics use case*, [[2002.01427](#)].
- 2933 [38] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022).
2934 <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoverview/>.

- 2935 [39] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*,
2936 *JINST* 11 (2016) p. P04008 [[1512.01094](#)].
- 2937 [40] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13 \text{ TeV}$*
2938 *using the ATLAS detector at the LHC*, *ATLAS-CONF-2019-021* (2019).
- 2939 [41] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design*
2940 *Report, Volume 1*, ATLAS-TDR-14; CERN-LHCC-99-014, Geneva (1999).
2941 <https://cds.cern.ch/record/391176>.
- 2942 [42] ATLAS Collaboration, *Status of the ATLAS Experiment*, Tech. Rep., Geneva (2010).
2943 <https://cds.cern.ch/record/1237407>.
- 2944 [43] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design*
2945 *Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
2946 <https://cds.cern.ch/record/2285585>.
- 2947 [44] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design*
2948 *Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
2949 <https://cds.cern.ch/record/2257755>.
- 2950 [45] ATLAS Collaboration, *The inner detector*, (2022).
2951 <https://atlas.cern/Discover/Detector/Inner-Detector>.
- 2952 [46] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*,
2953 ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
2954 <https://cds.cern.ch/record/1291633>.
- 2955 [47] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,
2956 *JINST* 13 (2018) p. T05008 [[1803.00844](#)].
- 2957 [48] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor*
2958 *tracker*, *JINST* 9 (2014) p. P08009 [[1404.7473](#)].
- 2959 [49] ATLAS Collaboration, *Calorimeter*, (2022).
2960 <https://atlas.cern/Discover/Detector/Calorimeter>.
- 2961 [50] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*,
2962 ATLAS-TDR-3; CERN-LHCC-96-042, Geneva (1996).
2963 <https://cds.cern.ch/record/331062>.
- 2964 [51] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in*
2965 *proton–proton collision data at $\sqrt{s} = 13 \text{ TeV}$* , *Eur. Phys. J. C* 76 (2016) p. 292

- 2966 [1603.05598].
- 2967 [52] ATLAS Collaboration, *Muon spectrometer*, (2022).
2968 <https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- 2969 [53] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* 77 (2017) p. 317 [1611.09661].
- 2970 [54] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*,
2971 *ATL-SOFT-PUB-2021-001* (2021).
- 2972 [55] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 2973 [56] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense Environments*, *ATL-PHYS-PUB-2015-006* (2015).
- 2974 [57] 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].
- 2975 [58] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, *JINST* 9 (2014) p. P09009 [1406.7690].
- 2976 [59] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS Experiment*, Ph.D. Thesis, Innsbruck U., 2013.
2977 <https://cds.cern.ch/record/1625231>.
- 2978 [60] 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].
- 2979 [61] 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].
- 2980 [62] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, *ATL-PHYS-PUB-2017-011* (2017).
- 2981 [63] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* 04 (2008) p. 063 [0802.1189].
- 2982 [64] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur.Phys.J. C*72 (2012)
2983 p. 1896 [1111.6097].

- 2996 [65] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04**
2997 (2008) p. 005 [[0802.1188](#)].
- 2998 [66] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS*
2999 *detector*, [ATLAS-CONF-2014-018](#) (2014).
- 3000 [67] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*
3001 *performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- 3002 [68] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*
3003 *the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 3004 [69] ATLAS Collaboration, *Jet energy scale measurements and their systematic*
3005 *uncertainties in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*,
3006 *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].
- 3007 [70] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a*
3008 *New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008)
3009 p. 242001 [[0802.2470](#)].
- 3010 [71] ATLAS collaboration, *Measurement of the associated production of a Higgs boson*
3011 *decaying into b-quarks with a vector boson at high transverse momentum in pp*
3012 *collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
3013 p. 136204 [[2008.02508](#)].
- 3014 [72] ATLAS collaboration, *Performance of jet substructure techniques for large- R jets in*
3015 *proton-proton collisions at $\sqrt{s} = 7\text{ TeV}$ using the ATLAS detector*, *JHEP* **09** (2013)
3016 p. 076 [[1306.4945](#)].
- 3017 [73] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in*
3018 *13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019)
3019 p. 135 [[1807.09477](#)].
- 3020 [74] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R* , *JHEP* **06** (2009) p. 059
3021 [[0903.0392](#)].
- 3022 [75] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
3023 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, [ATL-PHYS-PUB-2017-010](#)
3024 (2017).
- 3025 [76] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the*
3026 *Gaussian Sum Filter-based model for bremsstrahlung*, [ATLAS-CONF-2012-047](#) (2012).

- 3027 [77] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*
3028 *using the 2015 LHC proton–proton collision data*, **ATLAS-CONF-2016-024** (2016).
- 3029 [78] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
3030 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* , *Eur. Phys. J. C* **78** (2018) p. 903 [[1802.08168](#)].
- 3032 [79] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
3033 *ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].
- 3034 [80] ATLAS Collaboration, *Observation of Higgs boson production in association with a*
3035 *top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173
3036 [[1806.00425](#)].
- 3037 [81] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*
3038 *using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, *JHEP* **03**
3039 (2020) p. 145 [[1910.08447](#)].
- 3040 [82] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000)
3041 pp. 577–606 [[hep-ph/9912292](#)].
- 3042 [83] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**
3043 (2018) p. 030001.
- 3044 [84] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom*
3045 *and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets*,
3046 **ATL-PHYS-PUB-2014-008** (2014).
- 3047 [85] Nazar Bartosik, *Diagram showing the common principle of identification of jets*
3048 *initiated by b -hadron decays*, (2022).
3049 https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png.
- 3050 [86] ATLAS Collaboration, *Identification of Jets Containing b -Hadrons with Recurrent*
3051 *Neural Networks at the ATLAS Experiment*, **ATL-PHYS-PUB-2017-003** (2017).
- 3052 [87] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter*
3053 *Flavour Tagging in ATLAS*, **ATL-PHYS-PUB-2020-014** (2020).
- 3054 [88] ATLAS Collaboration, *Expected performance of the ATLAS b -tagging algorithms in*
3055 *Run-2*, **ATL-PHYS-PUB-2015-022** (2015).
- 3056 [89] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*
3057 *b -tagging algorithms for the 2017-18 LHC run*, **ATL-PHYS-PUB-2017-013** (2017).

- 3058 [90] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp*
3059 *collision dataset*, [[2211.16345](#)].
- 3060 [91] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,
3061 *Journal of High Energy Physics* 2004 (2004) p. 040–040.
- 3062 [92] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*
3063 *carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* 2007 (2007)
3064 p. 126–126.
- 3065 [93] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*
3066 *shower simulations: the powheg method*, *Journal of High Energy Physics* 2007 (2007)
3067 p. 070–070.
- 3068 [94] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*
3069 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*
3070 *Physics* 2010 (2010) .
- 3071 [95] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* 04 (2015)
3072 p. 040 [[1410.8849](#)].
- 3073 [96] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
3074 *ATL-PHYS-PUB-2016-020* (2016).
- 3075 [97] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An*
3076 *introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* 191 (2015) p. 159
3077 [[1410.3012](#)].
- 3078 [98] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,
3079 *ATL-PHYS-PUB-2014-021* (2014).
- 3080 [99] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* 867 (2013)
3081 p. 244 [[1207.1303](#)].
- 3082 [100] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*
3083 462 (2001) p. 152.
- 3084 [101] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* 70
3085 (2010) p. 823 [[1005.4568](#)].
- 3086 [102] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.*
3087 *Instrum. Meth. A* 506 (2003) p. 250.
- 3088 [103] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).

- 3089 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLLOTS/IDTR-2022-03/>.
- 3090 [104] S. Adorni Braccesi Chiassi, *Vector Boson Tagging in the Context of the Search for
3091 New Physics in the Fully Hadronic Final State*, Ph.D. Thesis, Geneva U., 2021.
3092 [10.13097/archive-ouverte/unige:158558](https://archive-ouverte.unige:158558).
- 3093 [105] E. Boos et al., *Generic user process interface for event generators*, [[hep-ph/0109068](https://arxiv.org/abs/hep-ph/0109068)].
- 3094 [106] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A
3095 standard format for Les Houches Event Files*, *Computer Physics Communications*
3096 **176** (2007) pp. 300–304.
- 3097 [107] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are
3098 universal approximators*, *Neural Networks* **2** (1989) pp. 359–366.
- 3099 [108] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by
3100 back-propagating errors*, *nature* **323** (1986) pp. 533–536.
- 3101 [109] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous
3102 activity*, *The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.
- 3103 [110] J. Hopfield, *Neural networks and physical systems with emergent collective
3104 computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the
3105 Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 3106 [111] A. F. Agarap, *Deep Learning using Rectified Linear Units (ReLU)*, *arXiv e-prints*
3107 (2018) p. [arXiv:1803.08375](https://arxiv.org/abs/1803.08375) [[1803.08375](https://arxiv.org/abs/1803.08375)].
- 3108 [112] S. Elfwing, E. Uchibe and K. Doya, *Sigmoid-Weighted Linear Units for Neural
3109 Network Function Approximation in Reinforcement Learning*, *arXiv e-prints* (2017)
3110 p. [arXiv:1702.03118](https://arxiv.org/abs/1702.03118) [[1702.03118](https://arxiv.org/abs/1702.03118)].
- 3111 [113] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural
3112 networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 3113 [114] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv
3114 e-prints* (2014) [[1412.6980](https://arxiv.org/abs/1412.6980)].
- 3115 [115] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,
3116 *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. [arXiv:2002.08772](https://arxiv.org/abs/2002.08772)
3117 [[2002.08772](https://arxiv.org/abs/2002.08772)].
- 3118 [116] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,
3119 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,

- 3120 *arXiv e-prints* (2018) p. arXiv:1806.01261 [[1806.01261](#)].
- 3121 [117] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,
3122 *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 3123 [118] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13$ TeV data*,
3124 **ATL-PHYS-PUB-2015-037** (2015).
- 3125 [119] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*
3126 *experiment using the 2015 and 2016 LHC proton–proton collision data at*
3127 $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 3128 [120] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*
3129 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, *arXiv e-prints* (2021)
3130 [[2007.08294](#)].
- 3131 [121] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,
3132 *arXiv e-prints* (2021) [[2105.14491](#)].
- 3133 [122] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al.,
3134 *Attention Is All You Need*, *arXiv e-prints* (2017) [[1706.03762](#)].
- 3135 [123] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*
3136 *e-prints* (2014) [[1412.6980](#)].
- 3137 [124] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).
3138 <https://github.com/onnx/onnx>.
- 3139 [125] ATLAS Collaboration, *Performance of Run 3 HLT b-tagging with fast tracking,*
3140 *Public b-Jet Trigger Plots for Collision Data*, (2022).
3141 [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/
3142 BJetTriggerPublicResults#Performance_of_Run_3_HLT_b_taggi](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults#Performance_of_Run_3_HLT_b_taggi).
- 3143 [126] ATLAS Collaboration, *Neural Network Jet Flavour Tagging with the Upgraded*
3144 *ATLAS Inner Tracker Detector at the High-Luminosity LHC*,
3145 **ATL-PHYS-PUB-2022-047** (2022).
- 3146 [127] ATLAS Collaboration, *Jet flavour tagging with GN1 and DL1d. generator*
3147 *dependence, run 2 and run 3 data agreement studies*, (2023).
3148 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/>.
- 3149 [128] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs*
3150 *Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].

- 3151 [129] ATLAS Collaboration, *Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$*
3152 *decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C*
3153 **81** (2021) p. 178 [[2007.02873](#)].
- 3154 [130] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson*
3155 *decaying into b -quarks with a vector boson at high transverse momentum in pp*
3156 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
3157 p. 136204 [[2008.02508](#)].
- 3158 [131] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated*
3159 *Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039
3160 [[1512.02572](#)].
- 3161 [132] ATLAS Collaboration, *Performance of jet substructure techniques for large- R jets in*
3162 *proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
3163 p. 076 [[1306.4945](#)].
- 3164 [133] ATLAS Collaboration, *Measurements of b -jet tagging efficiency with the ATLAS*
3165 *detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 3166 [134] ATLAS Collaboration, *Calibration of light-flavour b -jet mistagging rates using*
3167 *ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, **ATLAS-CONF-2018-006**
3168 (2018).
- 3169 [135] ATLAS Collaboration, *Measurement of b -tagging efficiency of c -jets in $t\bar{t}$ events using*
3170 *a likelihood approach with the ATLAS detector*, **ATLAS-CONF-2018-001** (2018).
- 3171 [136] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the*
3172 *ATLAS Detector*, **ATL-PHYS-PUB-2014-013** (2014).
- 3173 [137] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b -quark*
3174 *pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) p. 836
3175 [[1906.11005](#)].
- 3176 [138] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO*
3177 *calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010)
3178 p. 043 [[1002.2581](#)].
- 3179 [139] ATLAS collaboration, *Measurement of the Z/γ^* boson transverse momentum*
3180 *distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09**
3181 (2014) p. 145 [[1406.3660](#)].
- 3182 [140] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al.,

- 3183 *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* 72 (2012) p. 1889
3184 [[1111.2034](#)].
- 3185 [141] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*,
3186 *JHEP* 10 (2012) p. 155 [[1206.3572](#)].
- 3187 [142] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^\pm/HZ + 0$ and 1 jet at*
3188 *NLO with the POWHEG BOX interfaced to GoSam and their merging within*
3189 *MiNLO*, *JHEP* 10 (2013) p. 083 [[1306.2542](#)].
- 3190 [143] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to*
3191 *associated WH and ZH production at hadron colliders*, *Phys. Rev. D* 68 (2003)
3192 p. 073003 [[hep-ph/0306234](#)].
- 3193 [144] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the*
3194 *Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* 579 (2004) pp. 149–156
3195 [[hep-ph/0307206](#)].
- 3196 [145] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production*
3197 *at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.*
3198 107 (2011) p. 152003 [[1107.1164](#)].
- 3199 [146] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects*
3200 *in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* 72 (2012) p. 1868 [[1111.0761](#)].
- 3201 [147] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated*
3202 *WH production and decay at the LHC*, *JHEP* 04 (2014) p. 039 [[1312.1669](#)].
- 3203 [148] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron*
3204 *colliders: The fully differential NNLO QCD calculation*, *Phys. Lett. B* 740 (2015)
3205 pp. 51–55 [[1407.4747](#)].
- 3206 [149] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson*
3207 *at NNLO*, *JHEP* 06 (2016) p. 179 [[1601.00658](#)].
- 3208 [150] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,
3209 *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* 02 (2013) p. 078
3210 [[1211.5015](#)].
- 3211 [151] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via*
3212 *gluon fusion in the SM and the 2HDM*, *JHEP* 06 (2015) p. 065 [[1503.01656](#)].
- 3213 [152] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for*

- 3214 *gluon-induced Higgs Strahlung*, *JHEP* 11 (2014) p. 082 [[1410.0217](#)].
- 3215 [153] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron*
3216 *Collider in the 2-Higgs-doublet model*, *JHEP* 02 (2014) p. 023 [[1307.8122](#)].
- 3217 [154] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron*
3218 *colliders*, *Comput. Phys. Commun.* 184 (2013) pp. 998–1003 [[1210.5347](#)].
- 3219 [155] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte*
3220 *Carlo for heavy flavour hadroproduction*, *JHEP* 09 (2007) p. 126 [[0707.3088](#)].
- 3221 [156] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair*
3222 *cross-section at hadron colliders*, *Comput. Phys. Commun.* 185 (2014) p. 2930
3223 [[1112.5675](#)].
- 3224 [157] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with*
3225 *shower in POWHEG: s- and t-channel contributions*, *JHEP* 09 (2009) p. 111
3226 [[0907.4076](#)].
- 3227 [158] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top*
3228 *quark production*, *Phys. Rev. D* 81 (2010) p. 054028 [[1001.5034](#)].
- 3229 [159] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for*
3230 *t-channel single top quark production*, *Phys. Rev. D* 83 (2011) p. 091503 [[1103.2792](#)].
- 3231 [160] E. Re, *Single-top Wt-channel production matched with parton showers using the*
3232 *POWHEG method*, *Eur. Phys. J. C* 71 (2011) p. 1547 [[1009.2450](#)].
- 3233 [161] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated*
3234 *production with a W^- or H^-* , *Phys. Rev. D* 82 (2010) p. 054018 [[1005.4451](#)].
- 3235 [162] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al.,
3236 *Event generation with SHERPA 1.1*, *JHEP* 02 (2009) p. 007 [[0811.4622](#)].
- 3237 [163] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* 7 (2019) p. 034
3238 [[1905.09127](#)].
- 3239 [164] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
3240 *Phys. Rev. Lett.* 108 (2012) p. 111601 [[1111.5206](#)].
- 3241 [165] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* 12 (2008)
3242 p. 039 [[0808.3674](#)].
- 3243 [166] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour*

- 3244 *dipole factorisation*, *JHEP* 03 (2008) p. 038 [[0709.1027](#)].
- 3245 [167] S. Höche, F. Krauss, M. Schönher and F. Siegert, *QCD matrix elements + parton*
3246 *showers. The NLO case*, *JHEP* 04 (2013) p. 027 [[1207.5030](#)].
- 3247 [168] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson*
3248 *Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys.*
3249 *Rev. Lett.* 103 (2009) p. 082001 [[0903.2120](#)].
- 3250 [169] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* 43
3251 (2016) p. 023001 [[1510.03865](#)].
- 3252 [170] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to*
3253 *Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* 03
3254 (2012) p. 075 [[1112.5142](#)].
- 3255 [171] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo*
3256 *program for Higgs production in vector-boson fusion and Higgs strahlung at hadron*
3257 *colliders*, *Comput. Phys. Commun.* 195 (2015) pp. 161–171 [[1412.5390](#)].
- 3258 [172] ATLAS Collaboration, *Measurement of the Inelastic Proton–Proton Cross Section at*
3259 $\sqrt{s} = 13\text{ TeV}$ *with the ATLAS Detector at the LHC*, *Phys. Rev. Lett.* 117 (2016)
3260 p. 182002 [[1606.02625](#)].
- 3261 [173] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton*
3262 *distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C*75 (2015) p. 204
3263 [[1412.3989](#)].
- 3264 [174] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton*
3265 *distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.*
3266 D93 (2016) p. 033006 [[1506.07443](#)].
- 3267 [175] J. K. Anders and M. D’Onofrio, *V+Jets theoretical uncertainties estimation via a*
3268 *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
3269 <https://cds.cern.ch/record/2125718>.
- 3270 [176] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in*
3271 *association with a vector boson and decaying to a pair of b-quarks using large-R jets*,
3272 ATL-COM-PHYS-2019-1125 (2019).
- 3273 [177] J. Pumplin et al., *New generation of parton distributions with uncertainties from*
3274 *global qcd analysis*, *JHEP* 07 (2002) p. 012 [[0201195](#)].

- 3275 [178] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top*
3276 *hadroproduction in association with a W boson*, *JHEP* 07 (2008) p. 029 [[0805.3067](#)].
- 3277 [179] ATLAS Collaboration, *Measurement of VH, H → bb* production as a function of the
3278 *vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector*,
3279 *JHEP* 05 (2019) p. 141 [[1903.04618](#)].
- 3280 [180] ATLAS Collaboration, *Evidence for the H → bb* decay with the ATLAS detector,
3281 *JHEP* 12 (2017) p. 024 [[1708.03299](#)].
- 3282 [181] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino
3283 and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential*
3284 *Distributions*, *CERN-2012-002* (2012) [[1201.3084](#)].
- 3285 [182] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3.*
3286 *Higgs Properties*, *CERN-2013-004* (2013) [[1307.1347](#)].
- 3287 [183] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The*
3288 *roostats project*, *arXiv preprint arXiv:1009.1003* (2010) .
- 3289 [184] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, *Comput. Phys.*
3290 *Commun.* 77 (1993) pp. 219–228.
- 3291 [185] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for*
3292 *likelihood-based tests of new physics*, *Eur. Phys. J. C* 71 (2011) p. 1554 [[1007.1727](#)].
- 3293 [186] J. R. Andersen et al., *Les Houches 2015: Physics at TeV Colliders Standard Model*
3294 *Working Group Report*, [[1605.04692](#)].

3295 List of figures

<small>3296</small>	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.	<small>3297</small> 21
<small>3299</small>	2.2	Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [26]. 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.	<small>3300</small> 26
<small>3301</small>	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.	<small>3302</small> 26
<small>3303</small>	2.4	Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [26]. 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.	<small>3304</small> 27
<small>3305</small>	3.1	An overview of the CERN accelerator complex [36]. 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.	<small>3306</small> 30
<small>3307</small>	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. [37].	<small>3308</small> 31
<small>3309</small>			

3318	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. [38].	33
3324	3.4 Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [35]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.	35
3328	3.5 Average pile-up profiles measured by ATLAS during Run 2 [35].	35
3329	3.6 A 3D model of the entire ATLAS detector [42]. Cutouts to the centre of the detector reveal the different sub-detectors which are arranged in concentric layers around the nominal interaction point.	36
3332	3.7 A 3D model of the ATLAS ID showing the pixel, SCT and TRT sub- detectors [45].	37
3334	3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [38].	38
3336	3.9 A schematic cross-sectional view of the ATLAS IBL [46].	39
3337	3.10 The ATLAS calorimeters [49]. 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
3340	3.11 The ATLAS muon spectrometer [52].	43
3341	3.12 Particles (left) 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 recon- structed as a single merged cluster [57].	47
3345	3.13 A sketch of electron reconstruction using the ATLAS detector [77]. Electron reconstruction makes use of the entire ID and the calorimeters.	52
3347	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. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.	56

3350	3.15 Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [85]. 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.	56
3355	3.16 An overview of different high and low level taggers used in ATLAS. The low level taggers are IPxD, SV1 and JetFitter, and RNNIP and DIPS [61, 62, 86, 87]. The outputs of these taggers are fed into the high-level taggers DL1, DL1r and DL1d [89, 90].	58
3359	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
3365	4.2 b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [103]. 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. . . .	62
3372	4.3 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 pseudotrack represent the ideal performance of the ATLAS detector and are described in Section 3.4.1.	63
3379	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 $\sqrt{s} = 13\text{ TeV}$. Pseudotrack represent the ideal performance given the ATLAS detector, see Section 3.4.1.	65

3383	4.5	The average number of good (left) and wrong (right) IBL hits as a function of b -hadron p_T for tracks using baseline tracking (black) and the modified version of the outlier removal procedure (red).	67
3386	4.6	(left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified GX2F tracks. (right) The absolute value of the d_0 pull as a function of the truth b -hadron transverse momentum.	68
3389	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 θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result z is passed to an activation function. Each neuron can be thought of as a logistic regression model.	71
3393	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.	72
3396	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.	75
3400	5.4	The light-jet efficiency of the low level tagger SV1 for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$, 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.	76
3406	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.	79

3414	5.6 (left) Normalised histogram of the b -hadron track identification model 3415 output separated for tracks from the decay of a b -hadron and tracks from 3416 other sources. The two groups are further separated into those tracks which 3417 are fake. (right) The ROC curve for all tracks (solid line). The plots show 3418 tracks in the combined $t\bar{t}$ and Z' testing sample.	81
3419	5.7 The effect of applying the fake track identification algorithm together with 3420 the b -hadron decay track identification on the jet tagging performance of 3421 SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and 3422 $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) 3423 is compared to two working points of fake track removal, labelled “A” (red) 3424 and “B” (orange), which represent two different 2D working points of the 3425 track classification tools. Removal of fake tracks based on truth information 3426 is shown by the green curves.	82
3427	6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right). 3428 The existing approach utilises low-level algorithms (shown in blue), the 3429 outputs of which are fed into a high-level algorithm (DL1r). Instead 3430 of being used to guide the design of the manually optimised algorithms, 3431 additional truth information from the simulation is now being used as 3432 auxiliary training targets for GN1. The solid lines represent reconstructed 3433 information, whereas the dashed lines represent truth information [3]. . .	85
3434	6.2 The flow of information through a graph neural network layer for (left) a full 3435 implementation of the layer and (right) a deep sets model [117]. Reproduced 3436 from Ref. [116].	88
3437	6.3 The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of 3438 n_{tracks} , where each track is described by 21 track features ($n_{\text{tf}} = 21$). The 3439 jet features are copied for each of the tracks, and the combined jet-track 3440 vectors of length 23 form the inputs of GN1 [3].	93
3441	6.4 The network architecture of GN1. Inputs are fed into a per-track initialisation 3442 network, which outputs an initial latent representation of each track. 3443 These representations are then used to populate the node features of a 3444 fully connected graph network. After the graph network, the resulting node 3445 representations are used to predict the jet flavour, the track origins, and 3446 the track-pair vertex compatibility [3].	93

3447	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].	98
3448		
3449		
3450		
3451		
3452		
3453	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 $20 < p_T < 250$ 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].	101
3454		
3455		
3456		
3457		
3458		
3459		
3460	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 $250 < p_T < 5000$ 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].	101
3461		
3462		
3463		
3464		
3465		
3466		
3467	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].	102
3468		
3469		
3470		
3471		
3472		
3473	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].	104
3474		
3475		
3476		
3477		
3478		

3479	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].	104
3485	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].	106
3492	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].	106
3499	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.	108

3509	6.14 A schematic showing the truth track origin and vertex information (left)	
3510	compared with the predictions from GN1 (right). The diagrams show	
3511	the truth (left) and predicted (right) structure of a b -jet. The true and	
3512	predicted origins of the tracks is shown by the coloured circles along the	
3513	diagonal. The shaded black boxes show the groupings of tracks into different	
3514	vertices. Vertices reconstructed by SV1 and JetFitter are also marked. Class	
3515	probabilities p_b , p_c and p_u are rounded to three significant figures. GN1	
3516	improves over SV1 and JetFitter by successfully determining the origins	
3517	and vertex compatibility of all but one tracks in the jet.	108
3518	6.15 Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in	
3519	the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding	
3520	approaches. The exclusive vertexing approach includes single track vertices	
3521	while the inclusive approach requires vertices to contain at least two tracks.	
3522	Efficient vertex finding requires the recall of at least 65% of the tracks in	
3523	the truth vertex, and allows no more than 50% of the tacks to be included	
3524	incorrectly. Binomial error bands are denoted by the shaded regions. . .	111
3525	6.16 Heavy flavour vertex finding efficiency as a function of jet p_T for b -jets in	
3526	the Z' sample using the exclusive (left) and inclusive (right) vertex finding	
3527	approaches. The exclusive vertexing approach includes single track vertices	
3528	while the inclusive approach requires vertices to contain at least two tracks.	
3529	Efficient vertex finding requires the recall of at least 65% of the tracks in	
3530	the truth vertex, and allows no more than 50% of the tacks to be included	
3531	incorrectly. Binomial error bands are denoted by the shaded regions. . .	111
3532	6.17 Vertex finding efficiency for other secondary decays as a function of jet p_T	
3533	for light-jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right)	
3534	vertex finding approaches. The exclusive vertexing approach includes single	
3535	track vertices while the inclusive approach requires vertices to contain at	
3536	least two tracks. Efficient vertex finding requires the recall of at least 65%	
3537	of the tracks in the truth vertex, and allows no more than 50% of the tacks	
3538	to be included incorrectly. Binomial error bands are denoted by the shaded	
3539	regions.	112

3540	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.	112
3541		
3542		
3543		
3544		
3545		
3546		
3547	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].	115
3548		
3549		
3550		
3551		
3552		
3553		
3554	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. . .	117
3555		
3556		
3557		
3558		
3559	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. . .	117
3560		
3561		
3562		
3563		
3564	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 [125]. The ratio to the performance of the DL1d algorithm [90] 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.	118
3565		
3566		
3567		
3568		
3569		

3570	6.23 The c -jet rejection (left) and light-jet rejection (right) for the upgraded 3571 HL-LHC ATLAS detector with $\langle \mu \rangle = 200$ as a function of the b -jet efficiency 3572 for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of 3573 mass energy $\sqrt{s} = 14$ TeV [126]. The ratio to the performance of the DL1d 3574 algorithm is shown in the bottom panels. Binomial error bands are denoted 3575 by the shaded regions. The purple vertical dashed lines represent the most 3576 common working points used for b-tagging.	120
3577	6.24 The c -jet rejection (left) and light-jet rejection (right) for the upgraded 3578 HL-LHC ATLAS detector with $\langle \mu \rangle = 200$ as a function of the b -jet efficiency 3579 for jets in the Z' sample with $250 < p_T < 5000$ GeV for events with a centre of 3580 mass energy $\sqrt{s} = 14$ TeV [126]. The ratio to the performance of the 3581 DL1d algorithm is shown in the bottom panels. Binomial error bands are 3582 denoted by the shaded regions.	120
3583	7.1 Diagrams of the signal process (top) and $t\bar{t}$ background (bottom). The 3584 object to the right of centre are reconstructed within the large- R jet. For 3585 the $t\bar{t}$ background, the large- R jet contains a mistagged c -jet or (less often) 3586 a mistagged light-jet. The contribution in the 0-lepton channel results from 3587 hadronically decaying τ lepton, or a electron or muon which is out of the 3588 analysis acceptance. For the $t\bar{t}$ background process to be in the signal 3589 region, the b -jet on the left hand side must either be not within acceptance, 3590 not reconstructed, or not tagged.	130
3591	7.2 Leading large- R jet mass for the Z and $W+hf$ processes in the HP SR of 3592 the 0-lepton channel [176]. The renormalisation scale μ_r has been varied by 3593 a factor of 0.5 (1DOWN) and 2 (1UP). An exponential function is fitted to 3594 the ratio between the nominal and alternate samples.	142
3595	7.3 Leading large- R jet m_J inclusive in p_T^V for the $V+hf$ process modelled using 3596 both the SHERPA +PYTHIA8 (blue) and MADGRAPH +PYTHIA8 (red) 3597 samples [176].	143
3598	7.4 Leading large- R jet mass distribution for the combined WZ and ZZ pro- 3599 cesses, inclusive across all signal regions and lepton channels [176]. The 3600 renormalisation scale μ_R has been varied by a factor of 2 (UP) and 0.5 3601 (DOWN). The red and blue curves show the fitted results of the hyperbolic 3602 tangent function.	146

3603	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 [176]. 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.	146
3609	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 [130].	156
3619	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 (bot- tom) 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 [130].	157

3630	7.8 m_J distribution in data after subtraction of all backgrounds except for the	
3631	WZ and ZZ diboson processes. The contributions from all lepton channels	
3632	and signal regions are summed and weighted by their respective values of	
3633	the ratio of fitted Higgs boson signal and background yields. The expected	
3634	contribution of the associated WH and ZH production of a SM Higgs	
3635	boson with $m_H = 125$ GeV is shown scaled by the measured combined	
3636	signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to	
3637	its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and	
3638	systematic uncertainty is indicated by the hatched band. This error band is	
3639	computed from a full signal-plus-background fit including all the systematic	
3640	uncertainties defined in Section 7.3, except for the VH/VZ experimental	
3641	and theory uncertainties [130].	159
3642	7.9 Signal strength compatibility test between the (3+1)-POI fit (with the	
3643	three lepton channels fit separately) and the default (1+1)-POI fit. The	
3644	compatibility of the three channels is evaluated via a χ^2 difference test and	
3645	results in a p-value of 49% [130].	160
3646	7.10 Impact of systematic uncertainties on the fitted VH signal-strength pa-	
3647	rameter μ_{VH}^{bb} sorted in decreasing order. The boxes show the variations	
3648	of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual	
3649	nuisance parameter to its post-fit value modified upwards or downwards by	
3650	its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of	
3651	up- and down-variations can be distinguished via the dashed and plane box	
3652	fillings. The yellow boxes show the pre-fit impact (top x -axis) by varying	
3653	each nuisance parameter by ± 1 . The filled circles show the deviation of the	
3654	fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value	
3655	θ_0 expressed in standard deviations with respect to their nominal uncer-	
3656	tainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainities	
3657	on $\hat{\theta}$ with respect to their nominal uncertainities. The open circles show	
3658	the fitted values and uncertainities of the normalization parameters that are	
3659	freely floating in the fit. Pre-fit, these parameters have a value of one [176].	161
3660	7.11 Measured VH simplified template cross sections times the $H \rightarrow b\bar{b}$ and	
3661	$V \rightarrow$ leptons branching fractions in the medium and high $p_T^{V,t}$ bins [130].	164

3662 List of tables

<small>3663</small>	2.1 The fermions of the SM [16]. 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
<small>3664</small>		
<small>3665</small>		
<small>3666</small>	2.2 The bosons of the SM [16]. 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
<small>3667</small>		
<small>3668</small>		
<small>3669</small>		
<small>3670</small>	3.1 Overview of the different LHC runs [34, 35]. The average number of interac- tions 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 performance.	29
<small>3671</small>		
<small>3672</small>		
<small>3673</small>		
<small>3674</small>		
<small>3675</small>	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 PV and θ 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 [57]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.	46
<small>3676</small>		
<small>3677</small>		
<small>3678</small>		
<small>3679</small>		
<small>3680</small>		
<small>3681</small>		

3682	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 [57]. 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 Λ^0 decays, and hadronic interactions.	75
3689	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 [57], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.	77
3695	5.3 Hyperparameter for the track classification model	78
3696	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.	80
3698	5.5 Cut values for the fake and b -hadron decay track MVAs for the two defined working points. Working point “B” cuts more aggressively on the MVA 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).	81
3702	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 [57], 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].	90
3709	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 [111].	95

3713	6.3 The area under the ROC curves (AUC) for the track classification from 3714 GN1, compared to a standard multilayer perceptron (MLP) trained on a 3715 per-track basis. The unweighted mean AUC over the origin classes and 3716 weighted mean AUC (using as a weight the fraction of tracks from the 3717 given origin) is provided. GN1, which uses an architecture that allows track 3718 origins to be classified in a conditional manner as discussed in Section 6.4.3, 3719 outperforms the MLP model for both $t\bar{t}$ and Z' jets.	114
3720	7.1 Summary of the definitions of the different analysis regions . Signal enriched 3721 regions are marked with the label SR. There are regions with relatively large 3722 signal purity (HP SR) and with low purity (LP SR). Background enriched 3723 control regions are marked with the label CR. The shorthand ‘‘add’’ stands 3724 for additional small- R jets, i.e. number of small- R jets not matched to the 3725 Higgs-jet candidate. The medium and high p_T^V regions are referred to as 3726 Mp_T^V and Hp_T^V , respectively [130].	126
3727	7.2 Selections applied to baseline and signal electrons and muons. Signal leptons 3728 are additionally required to satisfy a $p_T > 27$ GeV (25 GeV in the 1-lepton 3729 channel).	127
3730	7.3 Event selection requirements for the boosted VH , $H \rightarrow b\bar{b}$ analysis channels 3731 and sub-channels [130]. Various channel dependent selections are used to 3732 maximise the signal efficiency and reduce the number of background events 3733 in each analysis region. The $\min[\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{small-}R \text{ jets})]$ selection is used to 3734 remove jets when the missing transverse momentum $\mathbf{E}_T^{\text{miss}}$ is pointing in the 3735 direction of the Higgs candidate, and the $\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T, \text{trk}})$ is used to reject 3736 events where the calorimeter missing transverse momentum $\mathbf{E}_T^{\text{miss}}$ is not 3737 pointing in the direction of the track-based missing transverse momentum 3738 $\mathbf{E}_{T, \text{trk}}^{\text{miss}}$. The $\Delta y(V, H_{\text{cand}})$ quantifies the rapidity difference between the 3739 reconstructed vector boson and Higgs candidate (H_{cand}). $m_{\ell\ell}$ refers to the 3740 invariant mass of the two leptons, while $p_T^{\ell_1}$ and $p_T^{\ell_2}$ refer to the p_T of 3741 the first and second leptons, respectively.	129

3742	7.4 Signal and background processes with the corresponding generators used for the nominal samples [130]. If not specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant (α_s). (★) The events were generated using the first PDF in the NNPDF3.0NLO set and subsequently reweighted to the PDF4LHC15NLO set [169] 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 [170, 171].	132
3754	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.	137
3757	7.6 0-lepton $W+jets$ nominal sample flavour composition and total event yield [176]. 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).	138
3760	7.7 1-lepton $W+jets$ nominal sample flavour composition and total event yield [176]. 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).	138
3763	7.8 0-lepton $Z+jets$ nominal sample flavour composition and total event yield [176]. 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
3766	7.9 2-lepton $Z+jets$ nominal sample flavour composition and total event yield [176]. 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
3769	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 [176].	141
3773	7.11 Diboson acceptance uncertainties. All uncertainties except channel extrapolation uncertainties are fully correlated between the ZZ and WZ processes and channels [176].	145

3776	7.12 Acceptance uncertainties for the $t\bar{t}$ background. The uncertainties on the 0- 3777 and 1-lepton channels are found to be similar, but are conservatively taken 3778 to be uncorrelated. The HP-to-LP SR uncertainty is applied in the HP SR. 3779 The Medium-to-high p_T^V uncertainty is applied in the high p_T^V bin, and is 3780 larger than the computed uncertainty in order to account for the known 3781 mismodelling of the $t\bar{t}$ p_T spectrum. [176].	147
3782	7.13 Single-top acceptance uncertainties. Uncertainties are derived in the 1-lepton 3783 channel but applied in a correlated fashion to both the 0- and 1-lepton 3784 channels. The HP-to-LP SR uncertainty is applied in the HP SR. The 3785 Medium-to-high p_T^V uncertainty is applied in the high p_T^V bin. The channel 3786 relative acceptance uncertainty is applied in the 0-lepton channel. [176]. .	148
3787	7.14 Factors applied to the nominal normalisations of the $t\bar{t}$, $W+hf$, and $Z+hf$ 3788 backgrounds, as obtained from the likelihood fit [130]. The errors represent 3789 the combined statistical and systematic uncertainties.	154
3790	7.15 Post-fit yields in the 0-lepton channel. Combined statistical and systematic 3791 uncertainties are shown [176].	154
3792	7.16 Post-fit yields in the 1-lepton channel. Combined statistical and systematic 3793 uncertainties are shown [176].	155
3794	7.17 Post-fit yields in the 2-lepton channel. Combined statistical and systematic 3795 uncertainties are shown [176].	155
3796	7.18 Breakdown of the absolute contributions to the uncertainty on the signal 3797 strength μ_{VH}^{bb} obtained from the (1+1)-POI fit. The average impact repre- 3798 sents the average between the positive and negative uncertainties on μ_{VH}^{bb} . 3799 The sum in quadrature of the systematic uncertainties attached to the cate- 3800 gories differs from the total systematic uncertainty due to correlations [176]. 3801	163