

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**

January 4, 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, potential improvements to, and the application of b -jet identification at the ATLAS experiment at the Large Hadron Collider (LHC). A focus is placed on the high transverse momentum regime, which is a critical region in which to study the Standard Model, but within which successful b -jet identification becomes difficult.

As b -jet identification 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 that may occur as the result of the decay of a long-flying b -hadron. The investigations revealed that the track quality selections applied during track reconstruction are suboptimal for tracks inside high transverse momentum b -jets, motivating future studies into the optimisation of these cuts.

To improve b -tagging performance two novel techniques are employed: the classification of the origin of tracks and the application of graph neural networks. An algorithm has been developed to classify the origin of tracks, which allows a more optimal collection of tracks to be selected for use in other algorithms. A graph neural network (GNN) jet flavour tagger has also been developed. This tagger requires only tracks and jet variables as inputs, making a break from previous algorithms, which relied on the outputs of several other algorithms. This model is trained to simultaneously predict the jet

flavour, track origins, and track-pair compatibility (i.e. vertexing), and demonstrates markedly improvements in b -tagging performance, both at low and high transverse momenta. The closely related task of c-jet identification also benefits from these methods.

Analysis of high transverse momentum $H \rightarrow b\bar{b}$ decays, where the Higgs boson is produced in association with a vector boson, was also performed using 139 fb^{-1} of 13 TeV proton-proton collision data from Run 2 of the LHC. The analysis is described, with a focus on the detailed modelling studies performed by the author. The analysis measured provided first measurements in this channel of the Higgs boson production in two high transverse momentum regions. The impact of applying the improved GNN-based b -tagging algorithms to the analysis is also studied.

Impact Statement

impact statement 500 words [link to ucl info](#)

Acknowledgements

Jonathan Shlomi for GNN.

Brian Moser and Francesco Di Bello provided useful guidance and insights into the studies in Chapter 7.

Conveners.

This thesis was made in L^AT_EX 2 _{ε} using the “heptesis” class [1].

Contents

1	Introduction	11
2	Theoretical Framework	13
2.1	The Standard Model	13
2.1.1	Quantum Electrodynamics	14
2.1.2	Quantum Chromodynamics	16
2.1.3	The Electroweak Sector	17
2.2	The Higgs Mechanism	18
2.2.1	Electroweak Symmetry Breaking	19
2.2.2	Fermionic Yukawa Coupling	22
2.2.3	Higgs Sector Phenomenology	24
3	The Large Hadron Collider and the ATLAS Detector	27
3.1	The Large Hadron Collider	28
3.2	Coordinate System & Collider Definitions	29
3.2.1	ATLAS Coordinate System	30
3.2.2	Track Parameterisation	31
3.2.3	Hadron Collider Definitions	31
3.3	The ATLAS Detector	35
3.3.1	Inner Detector	36
3.3.2	Calorimeters	39
3.3.3	Muon Spectrometer	41
3.3.4	The Trigger	42
3.4	Reconstructed Physics Objects	43
3.4.1	Tracks	43
3.4.2	Vertices	46
3.4.3	Jets	47
3.4.4	Leptons	49

3.4.5	Missing Transverse Momentum	51
4	Tracking and flavour tagging	52
4.1	<i>b</i> -hadron Reconstruction	53
4.1.1	Decay Topology	53
4.1.2	Challenges Facing <i>b</i> -hadron Reconstruction	55
4.2	Investigations into High p_T <i>b</i> -hadron Tracking	58
4.2.1	Shared Hits	59
4.2.2	Global χ^2 Fitter Outlier Removal	59
4.3	Conclusion	63
5	Track Classification MVA	64
5.1	Machine Learning Background	64
5.1.1	Neural Networks	65
5.1.2	Training with Gradient Descent	67
5.2	Track Truth Origin Labelling	68
5.3	Fake Track Identification Tool	69
5.3.1	Datasets	71
5.3.2	Model Inputs	72
5.3.3	Model Hyperparameters	72
5.3.4	Results	75
5.4	<i>b</i> -hadron Track Identification	76
5.5	Combined Approach	76
5.6	Conclusion	78
6	Graph Neural Network Flavour Tagger	80
6.1	Motivation	81
6.2	Graph Neural Network Theory	83
6.3	Experiemental Setup	84
6.3.1	Datasets	84
6.4	Model Architecture	85
6.4.1	Model Inputs	85
6.4.2	Auxiliary Training Objectives	87
6.4.3	Architecture	88
6.4.4	Training	92

6.5	Results	93
6.5.1	<i>b</i> -tagging Performance	94
6.5.2	<i>c</i> -tagging Performance	98
6.5.3	Ablations	99
6.5.4	Inclusion of Low-Level Vertexing Algorithms	101
6.5.5	Jet Display Diagrams	103
6.5.6	Vertexing Performance	105
6.5.7	Track Classification Performance	107
6.5.8	Looser Track Selection	110
6.6	Implementations of GN1	111
6.7	Conclusion	111
7	Boosted VHbb Analysis	114
7.1	Analysis Overview	115
7.1.1	Data & Simulated Samples	116
7.1.2	Object Reconstruction	117
7.1.3	Selection Criteria	119
7.1.4	Control Regions	121
7.1.5	Background Composition	121
7.2	Systematic Uncertainties & Background Modelling	123
7.2.1	Sources of Systematic Uncertainties	124
7.2.2	Implementation of Variations	126
7.2.3	Vector Boson + Jets Modelling	127
7.2.4	Diboson Modelling	134
7.3	Statistical Treatment	136
7.3.1	Likelihood Function	137
7.3.2	Background Normalisations	139
7.3.3	Asimov Dataset & Expected Results	140
7.4	Results	140
7.4.1	Post-fit Distributions	141
7.4.2	Observed Signal Strength & Significance	141
7.4.3	Impact of Systematics	145
7.5	Conclusion	149
8	Conclusion	150

Bibliography	151
---------------------	------------

List of figures	165
------------------------	------------

List of tables	176
-----------------------	------------

² Chapter 1

³ Introduction

- ⁴ This thesis describes various efforts in improving the understanding of the Higgs boson and its coupling to heavy flavour quarks, primarily through the improvement of the algorithms used to reconstruct and analyse jets.
- ⁷ Chapter 2 describes the theoretical foundations of the work presented in the rest of the thesis.
- ⁹ Chapter 3 describes the ATLAS detector at the CERN accelerator complex. Details of reconstructed physics objects are also provided.
- ¹¹ Chapter 4 provides an overview of tracking and *b*-tagging at ATLAS, and studies into the challenges of high transverse momentum *b*-tagging.
- ¹³ Chapter 5 describes a tool to predict the origins of tracks. The tool is used to improve *b*-tagging performance by the identification and removal of fake tracks before their input to the *b*-tagging algorithms.
- ¹⁶ Chapter 6 introduces a novel monolithic approach to *b*-tagging using graph neural networks and auxiliary training objectives.
- ¹⁸ Chapter 7 describes the measurement of the associated production of a Higgs boson decaying into a pair of *b*-quarks at high transverse momentum.
- ²⁰ Chapter 8 contains some concluding remarks.

- 21 The author's contribution to the work presented in this thesis is as follows.
- 22 **Tracking:** The author was been an active member of the Cluster and Tracking
23 in Dense Environments group for the duration of their qualification task on the
24 understanding of tracking performance at high transverse momentum. The author
25 played a key role in software r22 validation studies for the tracking group, including
26 the validation of the quasi-stable particle interaction simulation and the radiation
27 damage Monte-Carlo simulation. The author helped design and improve several
28 tracking software frameworks, and contributed to heavy flavour tracking efficiency
29 studies in dense environments.
- 30 ***b*-tagging:** The author has been an active member of the Flavour Tagging group
31 since September 2014. The author played a key role in investigating the performance
32 of the low level taggers at high transverse momentum and led studies into the
33 labelling and classification of track origins. Based on work by J. Shlomi, the author
34 helped develop a new flavour tagging algorithm which offers a large performance
35 improvement with respect to the current state of the art. The author was the primary
36 editor of a public note associated with this work [2]. The author also played a key
37 role in software r22 validation studies for the Flavour Tagging group, including the
38 validation of the quasi-stable particle interaction simulation. The author maintains
39 and contributes to various software frameworks used in the Flavour Tagging group,
40 and contributes to group documentation.
- 41 **Higgs:** The author was an active member of the Boosted VHbb analysis group. The
42 author performed various studies deriving systematic uncertainties for the $V+jets$
43 and diboson backgrounds. The author also produced and maintained samples, ran fit
44 studies and cross checks, and gave the diboson unblinding approval talk to the Higgs
45 group. The author also contributed to the developement of the analysis software.

⁴⁶ **Chapter 2**

⁴⁷ **Theoretical Framework**

⁴⁸ The Standard Model (SM) of particle physics is the theory describing all known
⁴⁹ elementary particles and their interactions via three of the four fundamental forces.
⁵⁰ Developed by merging the successful theories of quantum mechanics and relativity
⁵¹ in the second half of the 20th century, the SM's position today at the centre of our
⁵² understanding of the nature of the Universe is firmly established by an unparalleled
⁵³ level of agreement between the predictions from the model and experimental results
⁵⁴ [3, 4].

⁵⁵ The SM has predicted the discovery of the top and bottom quarks [5–7], the W
⁵⁶ and Z bosons [8], and the tau neutrino [9]. The last missing piece of the SM to be
⁵⁷ discovered was the Higgs boson, first theorised in the 1960s [10–12], and eventually
⁵⁸ observed at the LHC in 2012 [13, 14]. After its discovery, much ongoing work has
⁵⁹ been carried out performing detailed measurements of its mass and interactions with
⁶⁰ other particles.

⁶¹ In this chapter, an overview of the SM is given in Section 2.1, and a more detailed
⁶² discussion of the Higgs sector and Higgs phenomenology is provided in Section 2.2.

⁶³ **2.1 The Standard Model**

⁶⁴ The SM is formulated in the language of Quantum Field Theory (QFT). In this
⁶⁵ framework, particles are localised excitations of corresponding quantum fields, which
⁶⁶ are operator-valued distributions across spacetime.

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

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

The SM provides a mathematical description of how three of the four fundamental forces interact with the matter content of the Universe. The SM contains 12 spin-1/2 fermions, listed in Table 2.1, and five 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 [15]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.

78

2.1.1 Quantum Electrodynamics

Quantum electrodynamics (QED) is the relativistic quantum theory which describes the interaction between the photon and charged matter. Consider a Dirac spinor field $\psi = \psi(x)$ and its adjoint $\bar{\psi} = \psi^\dagger \gamma^0$, where ψ^\dagger denotes the Hermitian conjugate of ψ . The field ψ describes fermionic spin-1/2 particle, for example an electron. The

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 [15]. 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.

⁸⁴ Dirac Lagrangian density is

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

⁸⁵ 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

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

⁸⁸ Suppose some fundamental symmetry that requires invariance under a local $U(1)$
⁸⁹ gauge transformation

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

⁹⁰ where α varies over every spacetime point x . Under this transformation, the Dirac
⁹¹ equation transforms as

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

⁹² For the Dirac equation to remain invariant under the transformation in Eq. (2.4),
⁹³ a new field A_μ which transforms as $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x)$ must be added. The
⁹⁴ transformed interaction term

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

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

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

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

2.1.2 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the study of quarks, gluons and their interactions. Quarks and gluons carry colour charge, which comes in three kinds, called red, green and blue. While the $U(1)$ symmetry group in Section 2.1.1 was Abelian, the QCD Lagrangian is specified by requiring invariance under transformations from the non-Abelian $SU(3)$ group, making it a Yang-Mills theory [16] which requires the addition of self-interacting gauge fields. The infinitesimal $SU(3)$ group generators are given by $T_a = \lambda_a/2$, where λ_a are the eight Gell-Mann matrices. These span the space of infinitesimal group transformations and do not commute with each other, instead satisfying the commutation relation

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

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

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

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

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

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

124 The full QCD Lagrangian is then given by

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

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

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

135 2.1.3 The Electroweak Sector

136 The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam
137 (GWS) model of electroweak interaction [17–19]. The Lagrangian is specified by
138 requiring invariance under the symmetry group $SU(2)_L \otimes U(1)_Y$, as motivated by a
139 large amount of experimental data. Here, $SU(2)_L$ is referred to as weak isospin and
140 $U(1)_Y$ as weak hypercharge.

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

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

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

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

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

- 151 The weak force violates parity conservation [20–22], i.e. invariance under parity
152 transformations (mirror reflections). Only left handed fermions participate in the
153 weak interaction. Since there is no other force through which neutrinos interact with
154 other particles, there are no right handed neutrinos in the standard model.

155 2.2 The Higgs Mechanism

- 156 The Brout-Englert-Higgs mechanism (henceforth just “Higgs mechanism”) is the
157 mechanism through which the fundamental particles of the SM acquire mass [10–12].
158 Experimentally it was known that the weak force had a weak effective strength,
159 which was suggestive of a massive mediating gauge particle. However, directly
160 adding mass to the weak gauge bosons violates the non-Abelian symmetry of the
161 SM. Instead, the gauge bosons can gain mass through the interaction with a scalar
162 Higgs field which results from the spontaneous breakdown of symmetry as discussed
163 in Section 2.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as
164 discussed in Section 2.2.2. Section 2.2.3 described some basic phenomenology of the
165 Higgs particle relevant to hadron colliders.

¹⁶⁶ 2.2.1 Electroweak Symmetry Breaking

¹⁶⁷ Spontaneous symmetry breaking (SSB) is a key part of the Higgs mechanism. It
¹⁶⁸ is the transition of a physical system from a state of manifest symmetry to a state
¹⁶⁹ of hidden, or *broken*, symmetry. In particular, this applies to physical systems
¹⁷⁰ where the Lagrangian observes some symmetry, but the lowest energy vacuum states
¹⁷¹ do not exhibit that same symmetry. In other words, the symmetry is broken for
¹⁷² perturbations around the vacuum state.

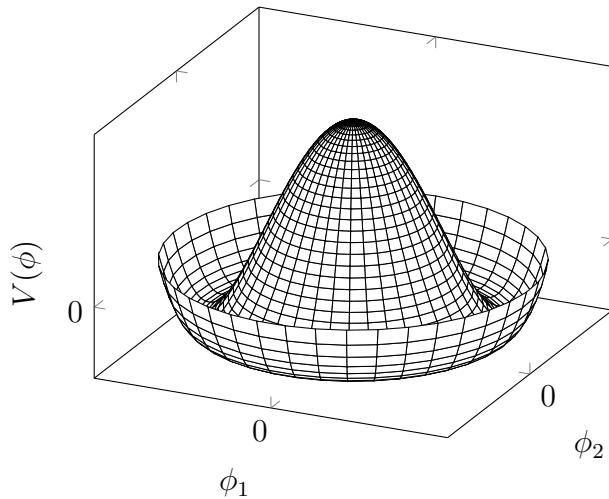


Figure 2.1: The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. The SM Higgs mechanism relies on a complex scalar doublet with a corresponding 5-dimensional potential.

¹⁷³ Consider gauge fields from the local $SU(2)_L \otimes U(1)_Y$ symmetry group discussed in
¹⁷⁴ Section 2.1.3 coupled to a complex scalar field $\phi = \phi(x)$. The scalar field ϕ transforms
¹⁷⁵ as a weak isospin doublet. Omitting the kinetic term of the gauge fields, and writing
¹⁷⁶ $\phi^2 \equiv \phi^\dagger \phi$, the Lagrangian is

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

¹⁷⁷ where the covariant derivative is given by

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

and T^a are the generators of $SU(2)$. The potential term $V(\phi)$ is made up of a quadratic and quartic term in the scalar field ϕ , which each contain an arbitrary parameter, respectively λ and μ . The quartic term gives the field self-interaction, and cannot be negative as this would lead to a potential that was unbounded from below. The quadratic term can be positive or negative. In the case where the quadratic term is positive, it is interpreted as a mass term for the scalar field. By choosing $\mu^2 < 0$ the field becomes unphysical due to its negative mass. In order to obtain a physical interpretation of the Lagrangian in Eq. (2.14) for the case where $\mu^2 < 0$, the field ϕ is expanded around the vacuum state. The vacuum expectation value (VEV) is the expected value of the field ϕ which minimises the potential $V(\phi)$ (equivalently the expected value of the field operator ϕ when the system is in a vacuum state, $|\langle \phi \rangle_0|^2 \equiv |\langle 0 | \phi | 0 \rangle|^2 \equiv \phi_0^2$). Minimising the potential gives a VEV of

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

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

Adding the particle content back to the theory by expanding the field around the vacuum state, and making a transformation to the unitary gauge to remove unphysical Nambu-Goldstone modes (which arise in the context of global symmetries [23, 24]), 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

204 in terms of the original fields A_μ^a and B_μ . This gives

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

205 where θ_W is the weak mixing angle defined by

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

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

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

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

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

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

225 Spontaneous symmetry breaking has modified the original symmetry group of the SM
226 $SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \otimes U(1)_Q$. Three broken generators from the symmetry
227 group $SU(2)_L \times U(1)_Y$ have been absorbed into the definition of the physical weak

228 vector bosons, giving them mass. The same methodology can be used to generate
229 the fermion masses, as shown in the next section.

230 2.2.2 Fermionic Yukawa Coupling

231 Adding the masses of the fermions by hand breaks the gauge invariance of the
232 theory. Instead, we can use a Yukawa coupling between the fermion fields and the
233 Higgs field in order to generate mass terms after spontaneous electroweak symmetry
234 breakdown [18]. In this way, the fermion masses are determined by both the respective
235 couplings to the Higgs field and the VEV of the Higgs field itself, which sets the
236 basic mass scale of the theory.

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

240 Lepton Masses

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

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

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

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

247 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
248 the first term in Eq. (2.22) which is quadratic in the electron field, and can therefore
249 be identified as the electron mass term. An interaction term between the electron
250 field e and the true vacuum Higgs field H is also present. Mass is generated for the
251 other lepton generations in the same way.

252 **Quark Masses**

253 The down-type quarks acquire their mass analogous to the leptons, with $\psi_L = (u_L, d_L)$
254 and $\psi_R = d_R$ for the first quark generation. Mass is generated for the up-type quarks
255 using the conjugate field to ϕ which transforms under $SU(2)$ as a doublet with
256 $Y = -1/2$. The conjugate field $\tilde{\phi}$ is constructed as

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

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

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

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

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

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

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

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

2.2.3 Higgs Sector Phenomenology

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

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

Since the Higgs boson couples proportional to mass as already mentioned, decays to heavier particles are favoured. The branching ratios of different Higgs boson decay modes are shown as a function of m_H in Fig. 2.4. Approximately 58% of the time the Higgs boson decays to a pair of b -quarks, the dominant decay mode. The next heaviest fermions are the tau lepton and the c -quark, decays to pairs of these particles happen approximately an order of magnitude less often. Decays to pairs of vector bosons are via a virtual off shell Higgs boson only. While the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ$ branching ratios are small compared with fermionic decay modes (around 0.2% for $H \rightarrow \gamma\gamma$), these decay channels were instrumental in the initial discovery of the Higgs due to the low level of background processes which mimic the final state.

This thesis presents a measurement of the Higgs bosons production rate using events with a Higgs boson produced in association with vector boson and decaying to a pair

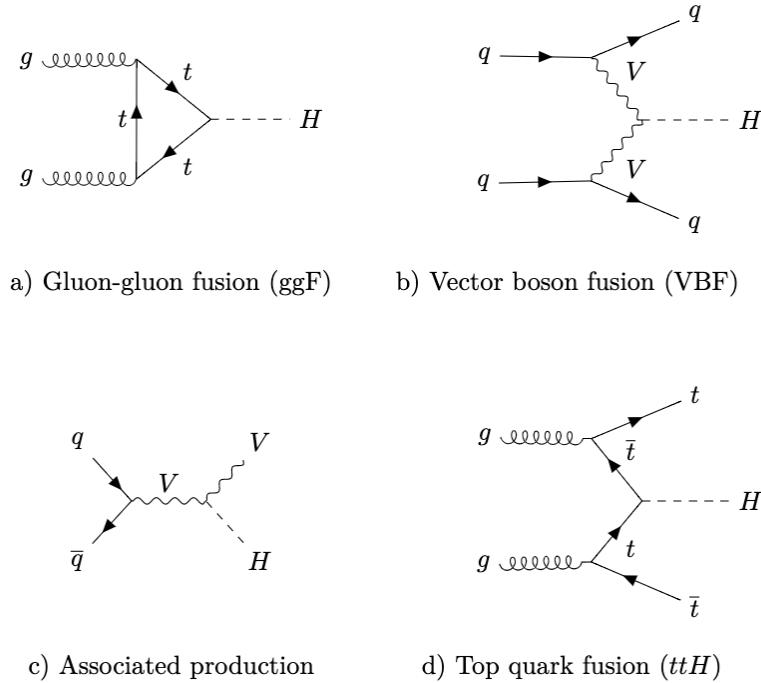


Figure 2.2: Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV.

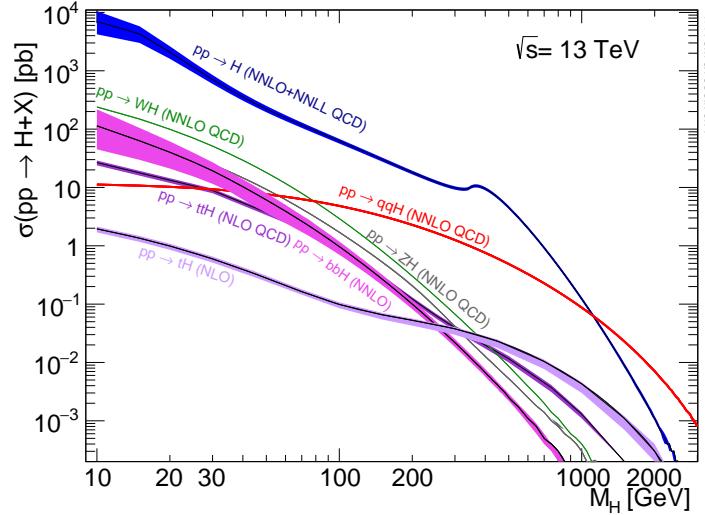


Figure 2.3: Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [25]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, Higgs boson production is dominated by gluon-gluon fussion, vector boson fusion, associated production with vector bosons, and top quark fusion.

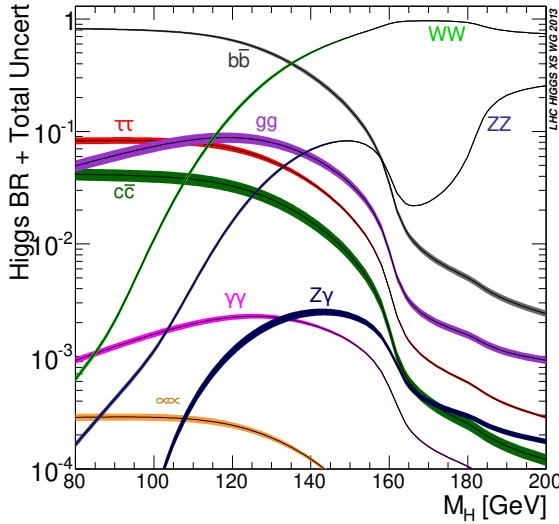


Figure 2.4: Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13 \text{ TeV}$ [25]. Uncertainties are shown in the shaded bands. At $m_H = 125 \text{ GeV}$, the Higgs predominantly decays to a pair of b -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of W bosons.

of b -quarks, i.e. $pp \rightarrow VH(bb)$. The $H \rightarrow b\bar{b}$ decay mode directly probes the Higgs coupling to the second generation fermions, and more specifically to the bottom quark. This coupling was first observed in 2018 [26, 27]. Ongoing work measuring the coupling strengths, in particular in the high energy regime, is the focus of the analysis presented in this thesis in Chapter 7.

₃₁₀ Chapter 3

₃₁₁ The Large Hadron Collider and the
₃₁₂ ATLAS Detector

₃₁₃ Since the completion of its construction in 2008, the Large Hadron Collider (LHC) [28]
₃₁₄ at CERN has extended the frontiers of particle physics through its unprecedented
₃₁₅ energy and luminosity. The LHC accelerates protons around a 27 km ring until they
₃₁₆ are travelling just 3 m s^{-1} slower than the speed of light, at which point they
₃₁₇ are made to collide. The protons travel round the ring 11,000 times per second in
₃₁₈ two concentric beams, which are guided by superconducting magnets cooled using
₃₁₉ liquid helium to -271.3°C (1.9 K). The beams travel in opposite directions around
₃₂₀ the ring and are crossed at four locations so that collisions between protons can
₃₂₁ take place. Around these collision points four specialised detectors, ALICE [29],
₃₂₂ CMS [30], LHCb [31] and ATLAS [32], are located to capture information about the
₃₂₃ products of the collisions.

₃₂₄ In this chapter, a brief overview of the LHC and the accelerator complex at CERN
₃₂₅ is given in Section 3.1. The coordinate system used at the ATLAS detector and
₃₂₆ other common definitions are introduced in Section 3.2. Next, an overview of the
₃₂₇ different detector systems is provided in Section 3.3, and finally descriptions of
₃₂₈ various commonly used reconstructed objects is given in Section 3.4.

³²⁹ 3.1 The Large Hadron Collider

³³⁰ The LHC is operated in multi-year *runs* during which beams of protons are circulated
³³¹ and collided. Between runs there are periods of shutdown while the accelerator and
³³² detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC
³³³ collided proton bunches, each containing more than 10^{11} particles, 20 million times
³³⁴ per second, providing 7 TeV proton-proton collisions at instantaneous luminosities
³³⁵ of up to $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The centre-of-mass energy was increased to 8 TeV
³³⁶ towards the end of Run 1 in 2012. Run 2, which spanned in 2015–2018, further
³³⁷ increased the the proton-proton collision energy to 13 TeV. During Run 2 the bunch
³³⁸ spacing was reduced, leading to a collisison rate of 40 MHz. Over the course of
³³⁹ Run 2 a total usable integrated luminosity of 139 fb^{-1} was recorded. 2022 marked the
³⁴⁰ beginning of Run 3 which, with a higher center of mass energy and peak luminosity,
³⁴¹ is expected to culminate in the approximate tripling of the dataset size. A summary
³⁴² of key information about each run is listed in Table 3.1.

Period	Year	\sqrt{s} [TeV]	$\langle\mu\rangle$	Bunch spacing [ns]	Luminosity [$\text{cm}^{-2} \text{ s}^{-1}$]
Run 1	2010–2012	7–8	18	50	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 [33,34]. 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. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.

³⁴³ An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is
³⁴⁴ at the final stage of a chain of accelerators which incrementally step-up the energy
³⁴⁵ of incoming protons. The first accelerator is Linac4, a linear accelerator which
³⁴⁶ accelerates negative hydrogen ions to an energy of 160 MeV. Upon leaving Linac4,
³⁴⁷ the ions are stripped of both electrons and the resulting protons are fed into the
³⁴⁸ Proton Synchrotron Booster (PSB), which increases the energy of the protons to
³⁴⁹ 2 GeV. The protons leaving the PSB are passed to the Proton Synchrotron (PS),
³⁵⁰ which increases the energy to 26 GeV, and then from the PS to the Super Proton
³⁵¹ Synchrotron (SPS) which further increases the energy to 450 GeV. Finally, the proton

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

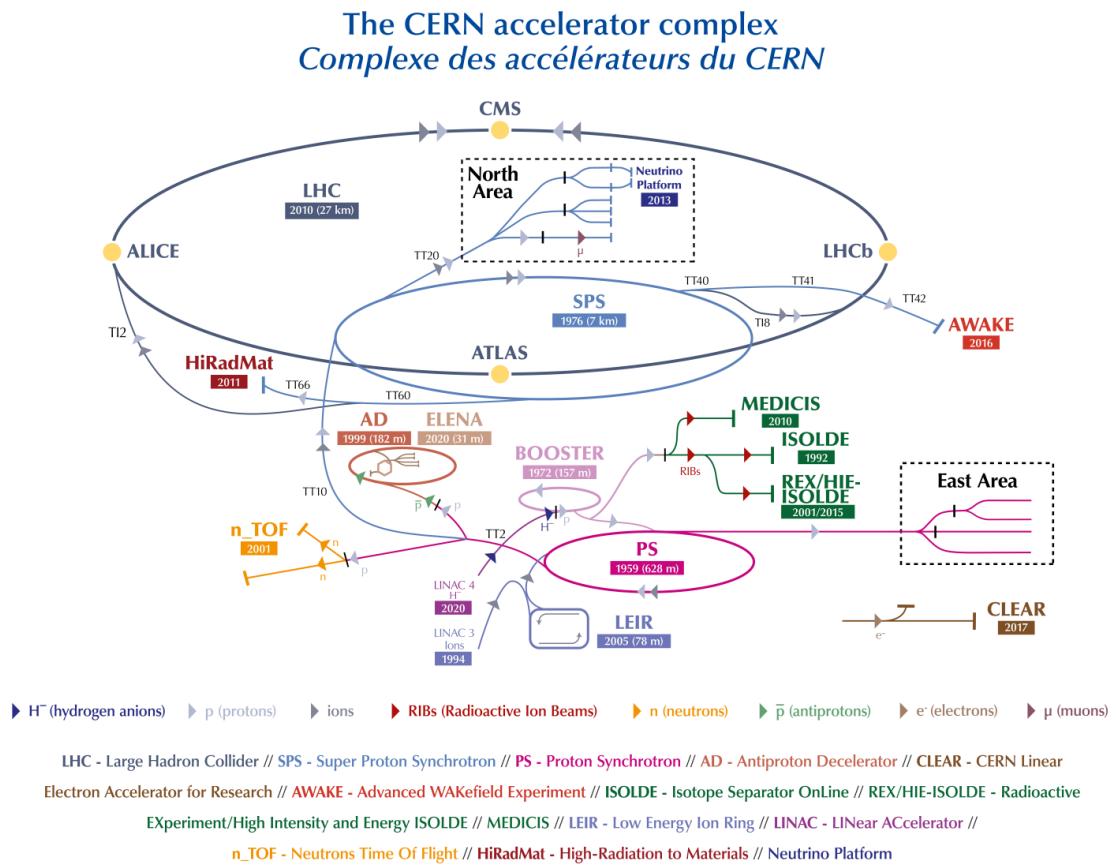


Figure 3.1: An overview of the CERN accelerator complex [35]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

³⁵⁴ 3.2 Coordinate System & Collider Definitions

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

359 3.2.1 ATLAS Coordinate System

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

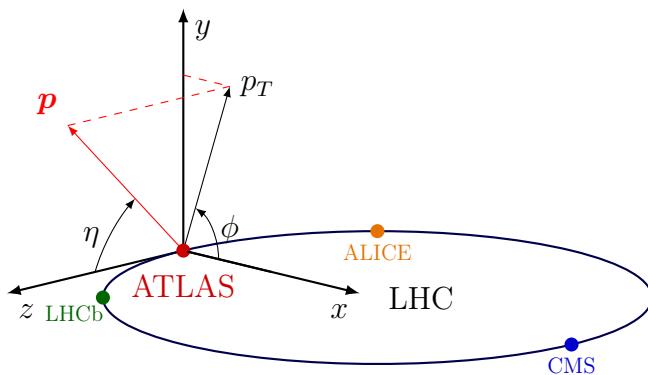


Figure 3.2: The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [36].

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

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

368 The pseudorapidity is a convenient quantity to work with as differences in η are
369 invariant under Lorentz boosts. In addition, particle production is constant as a
370 function of η .

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

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

³⁷³ Angular distance between two objects is measured in units of ΔR and is defined as
³⁷⁴ the sum in quadrature of the η and ϕ displacements

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

³⁷⁵ 3.2.2 Track Parameterisation

³⁷⁶ The trajectories of charged particle tracks are parameterised as a helix which is
³⁷⁷ fully specified using five parameters: $(d_0, z_0, \phi, \theta, q/p)$. Transverse and longitudinal
³⁷⁸ impact parameters (IP) d_0 and z_0 specify the closest approach of the trajectory of
³⁷⁹ a particle to the origin. The transverse IP d_0 and longitudinal IP z_0 are measured
³⁸⁰ with respect to the hard scatter primary vertex (see Section 3.4.2). ϕ and θ are
³⁸¹ the azimuthal and polar angles respectively, and q/p is the measured charge on the
³⁸² track¹ divided by the scalar 3-momentum. Fig. 3.3 shows each of these parameters
³⁸³ diagrammatically.

³⁸⁴ Impact parameter significances are defined as the IP divided by its corresponding
³⁸⁵ uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. When used in flavour tagging
³⁸⁶ (see Chapter 4), track IP significances are lifetime signed according to the track's
³⁸⁷ direction with respect to the jet axis and the primary vertex [38]. The sign IP
³⁸⁸ significances is positive if the track crosses the jet axis in front of the primary vertex
³⁸⁹ and negative if the crossing is behind the primary vertex.

³⁹⁰ 3.2.3 Hadron Collider Definitions

³⁹¹ Cross Section

³⁹² The cross section σ is closely related to the probability of an interaction between
³⁹³ two colliding particles, and is analogous to an effective cross-sectional area of the
³⁹⁴ particles. The cross section of a process depends on the transition matrix element,
³⁹⁵ obtained using the Feynman rules of the theory which are derived using QFT, and a
³⁹⁶ phase space integral. At hadron colliders such as the LHC, the proton-proton cross

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

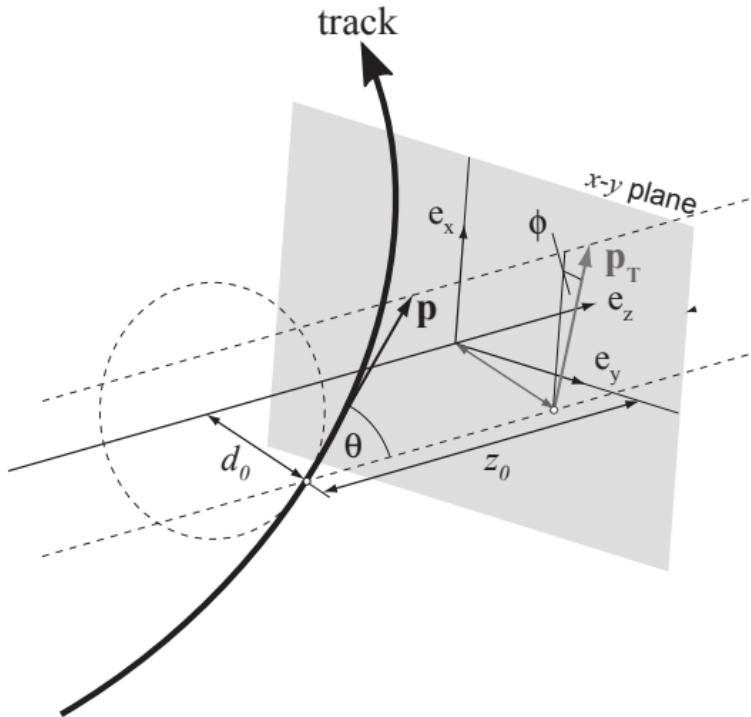


Figure 3.3: The track parameterisation used at the ATLAS detector. Five coordinates ($d_0, z_0, \phi, \theta, q/p$) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum \mathbf{p} and 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. [37].

³⁹⁷ section can be factorised as

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

³⁹⁸ The partonic cross section can be calculated at sufficiently high energies such as
³⁹⁹ those found at the LHC, while the parton distribution functions (PDFs) have to be
⁴⁰⁰ extracted from experimental results.

⁴⁰¹ Luminosity

⁴⁰² The total number of proton-proton collisions N is related to the total pp cross σ
⁴⁰³ section by the integrated luminosity L , as in

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

⁴⁰⁴ The instantaneous luminosity \mathcal{L} relates the cross section to the number of collisions
⁴⁰⁵ per unit time. For two colliding bunched proton beams, it is defined as

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

⁴⁰⁶ where n_1 and n_2 are the number of protons in the colliding bunches, f is the bunch
⁴⁰⁷ crossing frequency, and σ_x and σ_y are the rms width of the beam in the horizontal
⁴⁰⁸ and vertical directions.

⁴⁰⁹ The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In
⁴¹⁰ total, 139 fb^{-1} of usable physics data was collected over the three-year run. The
⁴¹¹ uncertainty on the total integrated luminosity is 1.7% [39].

⁴¹² Pile-up

⁴¹³ At the centre of the ATLAS detector, bunches of more than 10^{11} protons meet at a
⁴¹⁴ small crossing angle. Each bunch-crossing is called an *event*. There is generally at
⁴¹⁵ most one hard proton-proton scatter per event. Additional interactions are typically
⁴¹⁶ relatively soft and are known as *pile-up*. Pile-up from interactions within the same
⁴¹⁷ bunch-crossing is known as *in-time* pile-up while residual signatures from previous
⁴¹⁸ bunch-crossings is known as *out-of-time* pile-up. The number of pile-up interactions

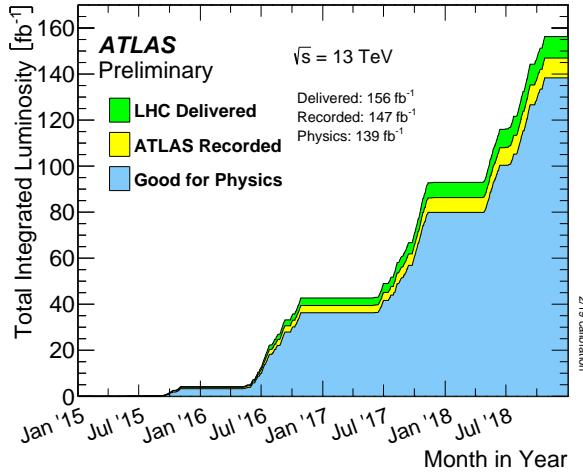


Figure 3.4: Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [34]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.

⁴¹⁹ is denoted μ , 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.

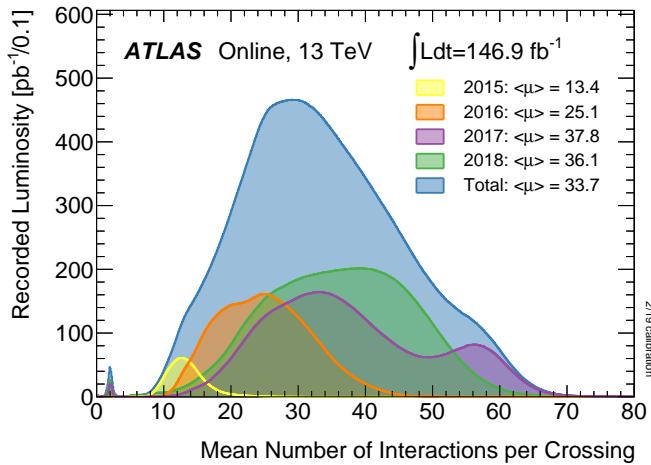


Figure 3.5: Average pile-up profiles measured by ATLAS during Run 2 [34]. Higher levels of pile-up are planned for Run 3.

421 3.3 The ATLAS Detector

422 The ATLAS² detector is made up of several specialised sub-detectors which are
423 arranged concentrically around the nominal interaction point at the centre of the
424 detector. The detector is designed to cover nearly the entire solid angle around the
425 collision point. In this section a condensed overview of each sub-detector is given, in
426 order of increasing radial distance from the point of collision. The inner tracking
427 detector is described in Section 3.3.1, the electromagnetic and hadronic calorimeters
428 in Section 3.3.2, the muon spectrometer in Section 3.3.3, and finally the trigger is
429 described in Section 3.3.4. More complete information on the detector can be found
430 in Ref. [32], while an overview of physics performance is given in [40].

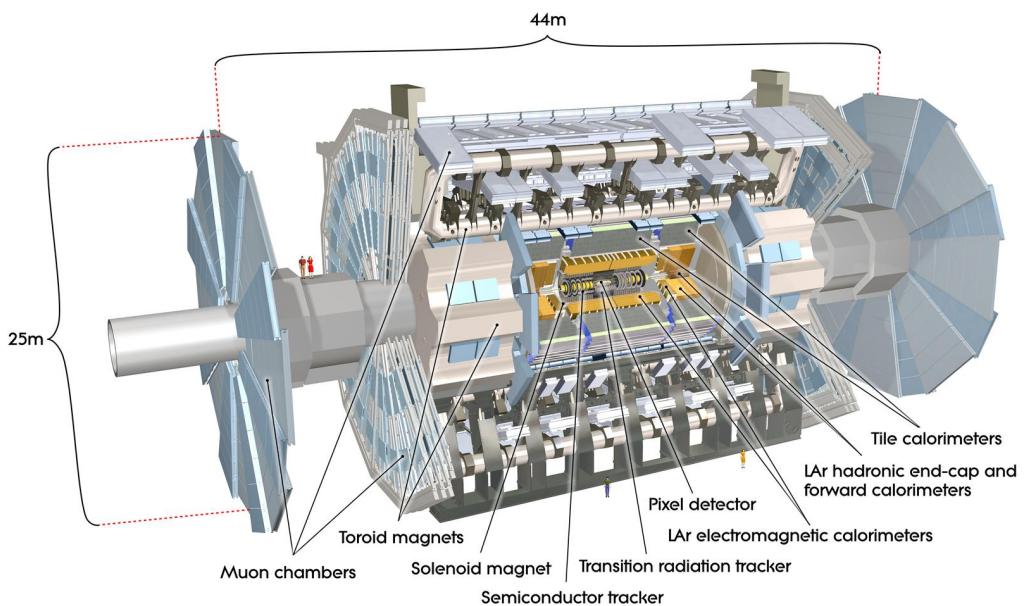


Figure 3.6: A 3D model of the entire ATLAS detector [41]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.

²A Toroidal LHC ApparatuS.

431 3.3.1 Inner Detector

432 The inner-detector system (ID) provides high-resolution charged particle trajectory
433 tracking in the range $|\eta| < 2.5$. The ID is immersed in a 2 T axial magnetic field,
434 produced by a superconducting solenoidal magnet, which enables the measurement
435 of particle momentum and charge. After Run 3, the ID will be replaced by the
436 ITk [42, 43].

437 The inner detector is made up of several sub-systems, shown in Figs. 3.7 and 3.8. The
438 high-granularity silicon pixel detector covers the vertex region and typically provides
439 four spacepoint measurements per track. It is followed by the silicon microstrip
440 tracker (SCT), which usually provides a further four spacepoint measurements per
441 track. These silicon detectors are complemented by the Transition Radiation Tracker
442 (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$.

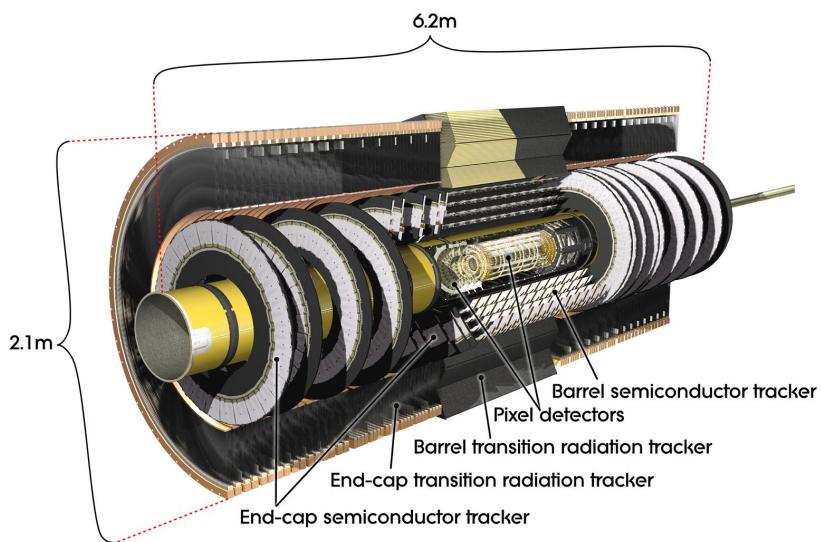


Figure 3.7: A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks [44].

443 The target inverse momentum resolution for the combined ID measurement is
444 parameterised as a function of the track transverse momentum and polar angle [40].
445 The parameterisation is given by

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

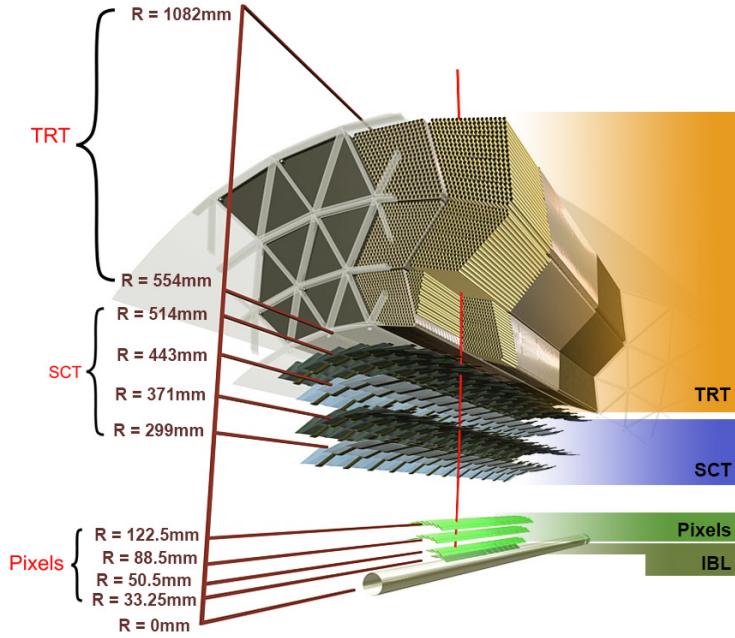


Figure 3.8: A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [37].

where \oplus denotes a sum in quadrature. For high- p_T tracks (e.g. $p_T \approx 100$ GeV) in the central region, $\sigma(1/p_T) \approx 0.4$ TeV $^{-1}$ corresponding to a relative error of 4%. The momentum resolution generally good enough to correctly identify the sign of the charge on particles up to the highest energies expected at the LHC. The transverse impact parameter resolution $\sigma(d_0)$ is parameterised similarly as

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

451 Pixel Detector

452 The silicon pixel detector is comprised of four cylindrical barrels at increasing radii
 453 from the beamline, and four disks on each side. The innermost barrel layer is
 454 the insertable B-layer (IBL), which was installed before Run 2 [45, 46] and lies
 455 approximately just 33 mm from the beam axis. The second-to-innermost layer is
 456 often referred to as the B-layer. The specification of the pixel detector determines the
 457 impact parameter resolution and the ability to reconstruct primary and secondary

vertices. The detector is required to have a high granularity (i.e. resolution) to maintain the low occupancy required to resolve nearby particles. Individual pixels are 50 μm in the transverse direction $R\phi$ and 400 μm in the longitudinal z direction (250 μm for the IBL). Cluster positions have a resolution of approximately 10 μm in $R\phi$ and 100 μm in z .

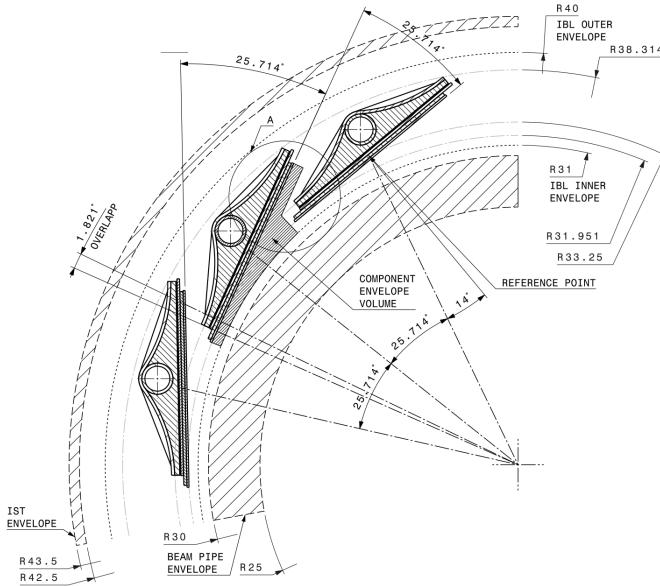


Figure 3.9: A schematic cross-sectional view of the ATLAS IBL [45].

463 Semi-Conductor Tracker (SCT)

464 The SCT is made up of four concentric barrel layers in the central region, and nine
 465 disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,
 466 with a small stereo angle (20 mrad) between the two layers enabling the z -coordinate
 467 to be measured from a pair of strip measurements. The SCT typically provides four
 468 precision spacepoint measurements (eight strip measurements) per track in the barrel
 469 region. These have intrinsic uncertainties of 17 μm in the transverse direction $R\phi$, and
 470 580 μm in the longitudinal direction z [47]. The measurements provide a contribution
 471 to the measurement of charged particle momentum and impact parameter, along
 472 with vertex position. Charge-particle tracks can be distinguished if separated by
 473 more than $\sim 200 \mu\text{m}$.

474 Transition Radiation Tracker (TRT)

475 The TRT is a straw-tube tracker which complements the higher-resolution silicon-
476 based tracks by offering a larger number of hits per track (typically around 30) and
477 a long lever arm, which aids the accurate measurement of particle momentum. It is
478 made up of approximately 300 000 drift tubes with a diameter of 4 mm which are filled
479 with an argon/xenon gas mixture. The walls of each tube are electrically charged,
480 and a thin conducting wire runs along the center. When a charged particle traverses
481 a tube, it ionises the gas and the resulting liberated electrons drift along the electric
482 field to the wire, where an associated charge is registered. In the barrel the straws
483 run parallel to the z -axis and therefore the TRT only provides tracking information
484 in $R\phi$. Straws are arranged radially in the end-caps. The resulting two-dimensional
485 spacepoints have a resolution of approximately 120 μm . The spaces between the
486 straws are filled with a polymer which encourages the emission of transition radiation,
487 aiding electron identification.

488 3.3.2 Calorimeters

489 The calorimeter system measures the energy of incident particles over the range
490 $|\eta| < 4.9$. There are two main sub-systems: the electromagnetic calorimeter (ECal),
491 which focuses on the measurement of electrons and photons, and the hadronic
492 calorimeter (HCal), which measures the energy of hadrons. Upon entering the
493 calorimeter, incident particles will interact with the detector material to produce a
494 shower of secondary particles with reduced energies. The charge deposited in this
495 process is measured to reconstruct the energy of the initial incident particle. The
496 two calorimeter sub-systems must provide strong containment of showering particles
497 to prevent punch-through of EM and non-muon particles to the HCal and muon
498 system respectively.

499 Liquid Argon (LAr) Electromagnetic Calorimeter

500 The more granular lead/liquid-argon ECal covers the region $|\eta| < 3.2$ and is split
501 into barrel (covering $|\eta| < 1.475$) and end-cap (covering $1.375 < |\eta| < 3.2$) regions.
502 EM calorimetry works by encouraging electrons and photons to interact with electri-

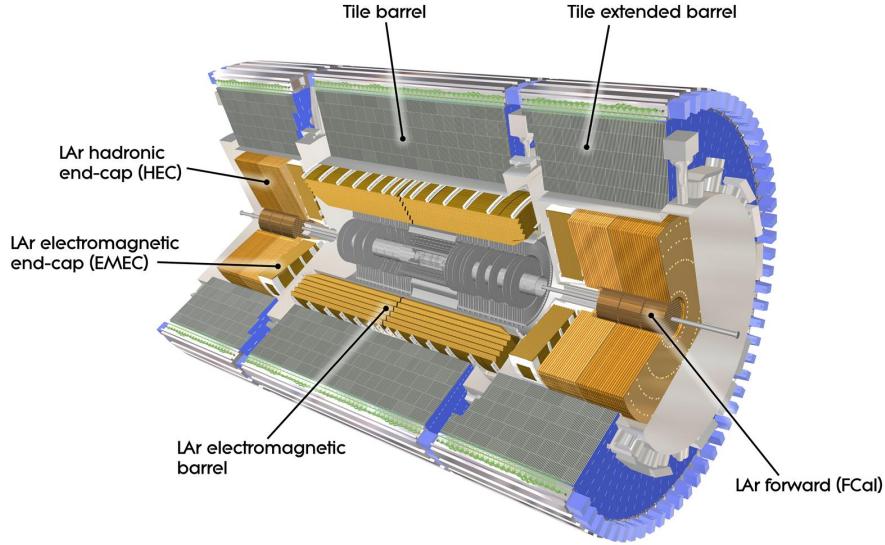


Figure 3.10: The ATLAS calorimeters [48]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic endcaps.

503 cally charged particles in detector material via bremsstrahlung ($e \rightarrow e\gamma$) and pair
 504 production ($\gamma \rightarrow e^+e^-$). The EM calorimeter uses lead absorber plates to initiate
 505 EM showers, resulting in secondary particles which ionise the surrounding liquid
 506 argon. The charge is collected on copper electrodes and read out. The accordion
 507 geometry of the ECal allows for a full coverage in ϕ without any azimuthal cracks.

508 The energy resolution of the LAr calorimeter is made up of a sampling and a constant
 509 term, which are summed in quadrature to produce the overall energy resolution. The
 510 sampling term contributes approximately $10\%/\sqrt{E}$, while the constant term adds an
 511 additional 0.7%. Photons with moderate transverse energy $E_T \approx 50 \text{ GeV}$ have an
 512 overall energy resolution of at most 1.6% over most of the pseudorapidity range. At
 513 lower $E_T \approx 10 \text{ GeV}$, the resolution is degraded to approximately 5%. The resolution
 514 measurements are obtained from test beam data [40].

515 Hadronic Tile Calorimeter

516 In the central barrel region with $|\eta| < 1.7$, the HCal uses a tile calorimeter with
 517 steel as an absorbing material, and scintillating tiles as the active material. Two

518 copper/liquid-argon calorimeter end-caps are also used. Incident hadrons interact
519 via the strong and electromagnetic forces with the absorber material, mainly loosing
520 energy due to multiple inelastic nuclear collisions. The active material captures the
521 resulting electrons and photons to measure the energy of the incident hadron.

522 The jet energy resolution of the HCal is parameterised as a function of the jet
523 transverse energy

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

524 corresponding to a jet energy resolution of 10% at a jet p_T of approximately 100 GeV
525 [49].

526 3.3.3 Muon Spectrometer

527 Due to their higher mass, muons easily pass unimpeded through the ID and calorime-
528 ters and therefore require specialised detectors for their measurement. The Muon
529 Spectrometer (MS) is made up of dedicated tracking and triggering hardware. The
530 precision tracking system uses three layers of monitored drift tubes with a barrel
531 region covering $|\eta| < 1.2$ and end-caps covering $1 < |\eta| < 2.7$. The inner layers of
532 the end-caps use cathode strip chambers to better cope with the high occupancy
533 in the forward region. Precision tracking resolution is approximately 50 μm . The
534 trigger system is comprised of resistive plate chambers in the barrel region covering
535 $|\eta| < 1.0$ and thin gap chambers in the end-cap regions covering $1 < |\eta| < 2.4$. A set
536 of three superconducting air-core toroidal magnets, each made up of eight coils, is
537 used in each of the barrel and end-caps to deflect the muons as they pass through
538 the MS, allowing their momentum and charge to be measured from the direction
539 and magnitude of curvature. The toroidal magnets generate a field which is largely
540 orthogonal to the muon trajectories which allows for maximum deflection. The
541 transverse momentum resolution has been measured to be approximately 1.7% in the
542 central region for low- p_T muons, increasing to 4% for high- p_T muons in the forward
543 regions [50].

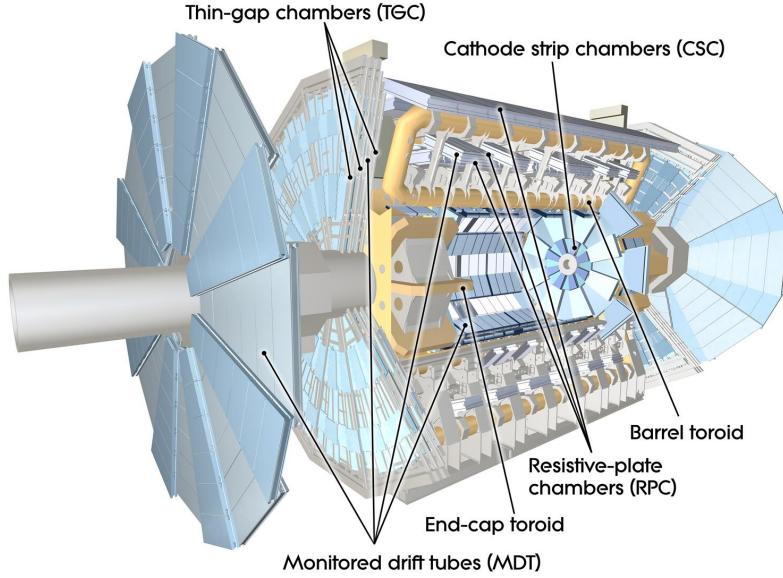


Figure 3.11: The ATLAS muon spectrometer [51].

544 3.3.4 The Trigger

545 The 25 ns bunch spacing used over the course of Run 2 corresponds to a bunch-
 546 crossing or event rate of 40 MHz (see Table 3.1). If the full information for the
 547 detector was written out for each event, this would correspond to the generation
 548 of 60 TB of data each second. This is more than can be feasibly read out from
 549 the hardware, processed and stored, requiring the use of a trigger system which
 550 quickly makes a decision about whether or not an event is potentially interesting
 551 and should be kept for further analysis. The trigger system is comprised of two
 552 levels which search for signs of electrons, muons, taus, photons, and jets, as well as
 553 events with large total or missing transverse energy. The hardware-based Level-1
 554 (L1) trigger uses coarse information from the calorimeters and MS to accept events at
 555 an average rate of 100 kHz approximately 2.5 μ s after the event. After the L1 trigger,
 556 the software-based High Level Trigger (HLT) makes use of 40 000 CPU cores to make
 557 a final selection on surviving events in approximately a few hundred milliseconds.
 558 The final event read-out rate is approximately 1.2 kHz, corresponding to 1.2 GB s^{-1}
 559 of permanent data storage. More information is provided in [52].

560 3.4 Reconstructed Physics Objects

561 Event reconstruction is the process of analysing the output from the detector to
562 determine the type and properties of particles present in an event. The reconstructed
563 event provides information about the underlying physics process that led to these
564 observable final state particles. Events passing the trigger selection (described in
565 Section 3.3.4) undergo offline reconstruction, which makes use of the full information
566 from the detector. Reconstruction and analysis of events relies on the extensive
567 ATLAS software stack, see Ref. [53] for more information.

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

570 3.4.1 Tracks

571 The reconstructed trajectories of charged particles are referred to as *tracks*. Track are
572 reconstructed from the energy depositions (called *hits*) left by the particles as they
573 traverse the the inner detector. Tracks are widely used for a variety of downstream
574 applications, including vertexing and jet tagging, so their accurate reconstruction
575 is a critical task. A comprehensive introduction to ATLAS tracking is available
576 in Ref. [54], while specific optimisations for dense environments are detailed in
577 Refs. [55, 56]. An overview of track reconstruction is given below.

578 Space-point Formation (Clustering)

579 When a charged particle traverses a silicon layer, charge can be collected in more
580 than one pixel or strip. This is due to the incident angle of the particles with respect
581 to the sensor, and also the drift of electrons between sensors caused by the magnetic
582 field. Clusters (also called *hits* or *space-points*) are formed by clustering neighbouring
583 pixels or strips and estimating locations of space-points using the shape and energy
584 distribution of the clusters.

585 Track Finding

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

602 Ambiguity Solving

603 Ambiguity solving was introduced as part of the ATLAS New Tracking effort [54],
604 which was intended to improve track reconstruction performance in dense envi-
605 ronments. In the ambiguity solver, track candidates are processed individually in
606 descending order of a track score. The track score quantifies the likelihood of the
607 track corresponding to the trajectory of a real particle. Scoring uses a number of
608 variables, including the number and positions of hits (preferring hits in more precise
609 regions of the detector), the transverse momentum of the track and the track fit
610 quality. The track fit quality describes the quality of the track as the χ^2 divided
611 by the number degrees of freedom on the track. A preference for high transverse
612 momentum tracks promotes the successful reconstruction of the more physically
613 interesting energetic particles, and suppresses the large number of wrong hits assigned
614 to low momentum tracks. The ambiguity solver also penalises tracks with missing
615 hits on the innermost detector layers.

616 During the processing of a given highest-scoring track candidate, the track is cleaned
617 (whereby problematic hits are removed), and, if the resulting track satisfies the quality
618 selection criteria, a high precision fit of the track parameters using the surviving hits
619 is performed. The high precision fit makes full use of all available information, and
620 uses an updated position and uncertainty estimate for each cluster obtained from
621 a Neural Network (NN) [57]. If the track has reached this stage without rejection
622 by passing various quality regiments, it is re-scored and returned to the list of track
623 candidates. If the same track is then processed again without requiring modification,
624 it is added to the final track collection. Track candidates that fall below a certain
625 quality cut are rejected. This selection does allow for the possibility of a track having
626 small number of shared hits.

list shared
hit cut?

627 Neural Network Cluster Splitting

628 As part of track cleaning, shared hits are classified by a NN to determine if they are
629 compatible with the characteristic features of a merged cluster [55, 57]. A merged
630 cluster is one made up of a combination of energy deposits from more than one
631 particle, which have become merged due to the closeness of the associated particles
632 and the limited resolution of the detector. While in general this event is rare, it
633 is common for clusters to become merged in dense environments, as discussed in
634 Section 4.1. If the cluster is predicted to be merged it is labelled as being freely
635 shareable, or *split*. Hits not compatible with the merged hypothesis can still be
636 shared by a limited number of tracks, but come with a penalty for the track which
637 may hinder its acceptance into the final track collection.

638 Pseudotracking

639 Pseudotracking uses Monte Carlo truth information to group together all the hits
640 left by each truth particle. Each collection of hits which, as a unit, satisfies basic
641 quality requirements is directly used in a full resolution track fit. If the track fit is
642 successful, a “pseudotrack” track is created and stored. If the track fit fails, or the
643 collection of hits does not pass the basic quality requirements (for example because
644 of a lack of hits) then the particle is said to be un-reconstructable. In this way,
645 pseudotracking performance represents the ideal reconstruction performance given the

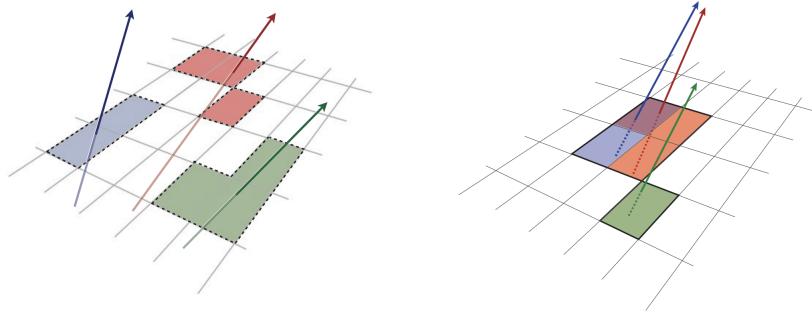


Figure 3.12: Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [56].

646 ATLAS detector, with perfect hit-to-track association and and track reconstruction
 647 efficiency. The approach was introduced in Ref. [58] as a way to obtain a fast
 648 approximation of tracking reconstruction for simulated data, however the technique
 649 has become a useful tool for studying tracking performance in general [55].

650 3.4.2 Vertices

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

659 Primary Vertices

660 Each proton-proton interaction happens at a *primary vertex* (PV). Primary vertices
 661 are iteratively reconstructed with tracks using the iterative vertex finder (IVF) [59].
 662 In Run 3, the IVF will be replaced with an adaptive multi-vertex finder (AMVF) [60].

663 The *hard scatter vertex* of an event is chosen as the primary vertex whose associated
664 tracks have the largest sum of transverse momentum squared, $\Sigma(p_T^2)$.

665 **Secondary Vertices**

666 Secondary vertices (SV) occur when a particle radiates or decays at a sufficient
667 distance from the primary vertex to be resolved from the primary vertex (see
668 Section 4.1.1). Two widely used secondary vertexing tools are used within ATLAS:
669 SV1 and JetFitter [61]. Each attempts to reconstruct secondary vertices inside a jet
670 using the tracks associated to that jet (see Section 3.4.3 for more information about
671 track association). SV1 by design attempts to reconstruct only a single inclusive
672 vertex per jet. This inclusive vertex groups all b -hadron decay products, including
673 tracks from the b -hadron decay itself and tracks from $b \rightarrow c$ decays. The second tool,
674 JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary
675 vertices from b -hadron decays are reconstructed separately to tertiary vertices from
676 $b \rightarrow c$ decay chains.

677 **3.4.3 Jets**

678 Jets are an aggregate reconstructed object corresponding to a collection of collimated
679 stable particles which results from a decay chain of an quark or gluon progenitor. Jets
680 are built by clustering constituent objects (e.g. tracks or calorimeter clusters) using
681 a jet finding algorithm, for example the anti- k_t algorithm [62], which is implemented
682 in FASTJET [63].

683 **EMTopo Jets**

684 EMTopo jets are reconstructed from noise-suppressed topological clusters (topoclus-
685 ters) of calorimeter energy depositions. The clustering uses the energy significance
686 of each cell, defined as

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

where E_{cell} is the energy measured in a given calorimeter cell, and $\sigma_{\text{noise}, \text{cell}}$ is the expected level of noise on the cell (e.g. from pile-up interactions). Topoclusters are formed from a seed cell with a large S_{cell} , and expanded by iteratively adding neighbouring cells with a sufficiently large energy significance. Collections of topoclusters are then clustered into a jet using the anti- k_t algorithm with a radius parameter of 0.4 (small- R jets) or 1.0 (large- R jets). More information is available in Ref. [64].

693 Particle Flow Jets

694 Particle-flow (PFlow) jets are reconstructed from particle-flow objects [65] using
 695 the anti- k_t algorithm with a radius parameter of 0.4. Particle-flow objects integrate
 696 information from both the ID and the calorimeters, improving the energy resolution
 697 at high transverse momenta and reducing pile-up contamination. The PFlow jet
 698 energy scale is calibrated according to Ref. [66].

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

704 Jet flavour labels are assigned according to the presence of a truth hadron within
 705 $\Delta R(\text{hadron}, \text{jet}) < 0.3$ of the jet axis. If a b -hadron is found the jet is labelled a b -jet.
 706 In the absence of a b -hadron, if a c -hadron is found the jet is called a c -jet. If no b -
 707 or c -hadrons are found, but a τ is found in the jet, it is labelled as a τ -jet, else it is
 708 labelled as a light-jet.

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

710 Large- R Jets

711 Large- R jets have a radius parameter $R = 1.0$ and are built by clustering topological
 712 calorimeter clusters using the anti- k_t algorithm [67]. The large radius parameter
 713 is especially useful for containing the decay products of a boosted Higgs boson, as
 714 discussed in Chapter 7. Due to their large size, large- R jets benefit from a grooming
 715 procedure called trimming which remove soft contaminants inside the jet [68, 69].

716 Trimming aims to remove jet constituents from pile-up and the underlying event,
 717 which helps to improve the jet mass resolution and its robustness to varying levels
 718 of pile-up. The jet mass is computed using a combination of information from the
 719 calorimeters and ID, and a calibration to data is applied [70].

720 Track-jets

721 Track-jets are built by clustering tracks using the anti- k_t clustering algorithm and
 722 are used in the analysis described in Chapter 7. The radius parameter is allowed
 723 to vary with transverse momentum such that a broader cone (up to $R = 0.4$) is
 724 used for low- p_T track-jets and a narrower cone (down to $R = 0.02$) for high- p_T
 725 track-jets [71, 72]. The narrower cone is better suited to clustering highly collimated
 726 jet constituents at high- p_T . Truth flavour labels for track-jets are derived using the
 727 same $\Delta R(\text{hadron}, \text{jet}) < 0.3$ matching scheme as used for PFlow jets.

728 3.4.4 Leptons

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

732 Electrons

733 Electron candidates are reconstructed by matching PV-compatible³ inner detector
 734 tracks to topological calorimeter clusters. The track-cluster matching criteria takes
 735 into account the significant energy loss of the electron due to bremsstrahlung. If a
 736 match is found, a refit of the track is performed using the Gaussian Sum Filter (GSF)
 737 [73], which better handles trajectory reconstruction in the presence of bremsstrahlung.
 738 Various identification criteria are then applied to the candidates using a likelihood-
 739 based (LH) method to improve purity. These include requirements on the track
 740 quality and cluster matching, the shape of electromagnetic shower in the ECal,
 741 leakage into the HCal, and the amount of transition radiation detected in the TRT.

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.

- 742 Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may also be applied. A full description can be obtained from Ref. [74].

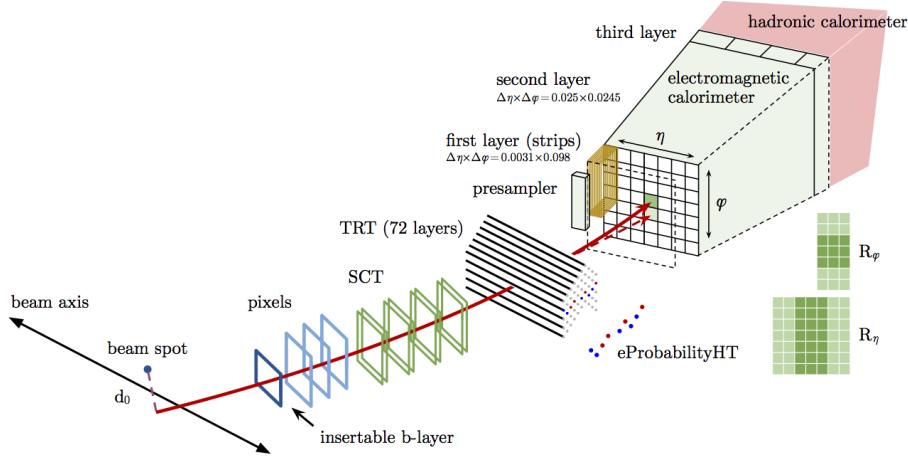


Figure 3.13: A sketch of electron reconstruction using the ATLAS detector [74]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.

743

744 Muons

745 Muon reconstruction makes use of the dedicated MS (see Section 3.3.3), the tracks 746 from the ID, and the presence of characteristic signatures in the calorimeters. Muon 747 tracks (i.e. a track reconstructed in the MS) are reconstructed by connecting straight-748 line track segments, which are identified via a Hough transform, and combined into 749 an approximately parabolic trajectory. Finally, a global χ^2 fit is performed, taking 750 into account possible interactions between the muon and the detector material. A 751 reconstructed muon is called *combined* if it completes successful matching to an 752 ID track. Combined muons undergo a further fit with the combined ID and MS 753 hits, with the energy loss due to the traversal of the calorimeters being taken into 754 account.

755 After reconstruction, candidate muons further undergo an identification processes 756 which helps to efficiently identify prompt muons whilst rejecting background sig-757 nals (e.g. non-prompt muons from pion and kaon decays, the punch-through of a 758 hadron from the calorimeter, or the semi-leptonic decay of a heavy flavour hadron). 759 Combined muon identification takes into account discrepancies in the p_T and charge

760 measurements in the MS and ID, and the χ^2 of the combined track fit. Selections
761 on the number of hits in the ID and MS are also applied. At the medium identifi-
762 cation working point, approximately 96% of muons with $20 \text{ GeV} < p_T < 100 \text{ GeV}$
763 are successfully identified. On top of the identification requirements, a number of
764 isolation requirements can also be applied to further suppress background signals. In
765 the region $|\eta| < 2.2$, the momentum resolution of reconstructed muons is 1.7%.

766 More information on muon reconstruction, identification and isolation can be found
767 in Ref. [75].

768 3.4.5 Missing Transverse Momentum

769 An imbalance in the final state transverse momentum can occur as a result of
770 incomplete measurement of the final state particles. In particular, neutrinos are
771 not measured by the detector and contribute to the missing transverse momentum
772 $\mathbf{E}_T^{\text{miss}}$. Incomplete detector acceptance and inaccuracies in the reconstruction of the
773 final state can also contribute to the missing transverse momentum of an event. In
774 order to calculate the missing transverse momentum, the negative vector sum of
775 the momentum of all photons, leptons and small- R jets with $p_T > 20 \text{ GeV}$ is taken.
776 The momenta of tracks associated to the primary vertex are also taken into account.
777 The magnitude of $\mathbf{E}_T^{\text{miss}}$ is written E_T^{miss} . More information about missing transverse
778 momentum reconstruction is provided in [76].

779

Chapter 4

780

Tracking and flavour tagging

781 Many ATLAS analyses rely on flavour tagging, which is the identification of jets
782 instantiated by heavy-flavour hadrons (b -hadrons and c -hadrons) as opposed to those
783 instantiated by light-flavour hadrons. In particular, b -tagging is the identification of
784 jets originating only from b -hadrons (i.e. b -jets). The b -jet identification algorithms
785 (also called *taggers*) work by identifying the unique signatures of b -jets, which are
786 outlined in Section 4.1. The various b -tagging algorithms ultimately take as their
787 input information about the reconstructed jet and its associated tracks. Successful
788 b -tagging relies therefore on the efficient and accurate reconstruction of tracks, and
789 especially those tracks corresponding to the products of b -hadron decays.

790 The current ATLAS flavour tagger, DL1r [77], is a deep neural network which
791 accepts as inputs the outputs of a number of independently optimised *low-level*
792 algorithms [61]. Correspondingly, DL1r is referred to as a *high-level* tagger (i.e. one
793 that uses a multivariate approach to combine the outputs of the low-level taggers).
794 Each of these low-level algorithms reconstructs a distinct feature of the experimental
795 signature of heavy flavour jets using the tracks associated to the jet. The low-level
796 algorithms are a combination of manually optimised reconstruction algorithms, for
797 example the SV1 and JetFitter algorithms that reconstruct displaced decay vertices,
798 and trained taggers such as RNNIP and DIPS that use the IP and hit information
799 from a variable number of tracks to identify the flavour of the jet [61, 78–80].

800 As the different b -tagging algorithms ultimately rely on tracks, accurate and efficient
801 track reconstruction is essential. This chapter summarises the challenges facing
802 tracking and b -tagging at high transverse momentum with an investigation into track

803 reconstruction performance in Section 4.1. Some preliminary investigations into
804 improving tracking in this regime are investigated in Section 4.2.

805 4.1 *b*-hadron Reconstruction

806 This section outlines the typical detector signature of a *b*-hadron in Section 4.1.1
807 and discusses some associated reconstruction difficulties in Section 4.1.2.

808 4.1.1 Decay Topology

809 *b*-hadrons are quasi-stable bound states of a bottom quark and one or more lighter
810 quarks. Collectively, these are the *B*-mesons (e.g. $B^+ = u\bar{b}$, $B^0 = d\bar{b}$) and baryons
811 (e.g. $\Lambda_b^0 = udb$). After a *b*-quark is produced as the result of a proton-proton collision,
812 they quickly hadronise. The hadronisation process is hard – around 70-80% of
813 the *b*-quark’s momentum is passed to the *b*-hadron, with the rest being radiated
814 as prompt hadronisation or fragmentation particles. See Ref. [81] for a more in
815 depth discussion on hadronisation and the closely related process of fragmentation.
816 Henceforth the combined hadronisation and fragmentation products will be referred
817 to collectively as fragmentation.

818 *b*-hadrons are interesting objects of study due to their relatively long proper lifetimes
819 $\tau \approx 1.5$ ps [82]. This lifetime corresponds to a proper decay length $c\tau \approx 450$ μm . In
820 the rest frame of the detector, the typical *b*-hadron travels a distance

$$d = \gamma\beta c\tau \approx \gamma c\tau \quad (4.1)$$

821 before decaying, where in the high energy limit $\gamma = E_b/m_b$ and $\beta = v/c = 1$.

822 For a 50 GeV *b*-hadron, this gives $d \approx 4.5$ mm, which is displaced enough to be
823 resolved from the primary vertex. Meanwhile for a 1 TeV *b*-hadron, $d \approx 90$ mm –
824 well beyond the radius of the first pixel layer (the IBL) which is situated at a radius
825 of approximately 33 mm from the center of the detector (the distance varies due
826 to the interleaved structure) Fig. 4.1 shows how the mean decay radius varies as a
827 function of *b*-hadron p_T . This significant displacement is characteristic of *b*-jets and
828 makes it possible to reconstruct secondary vertices at the *b*-hadron decay point.

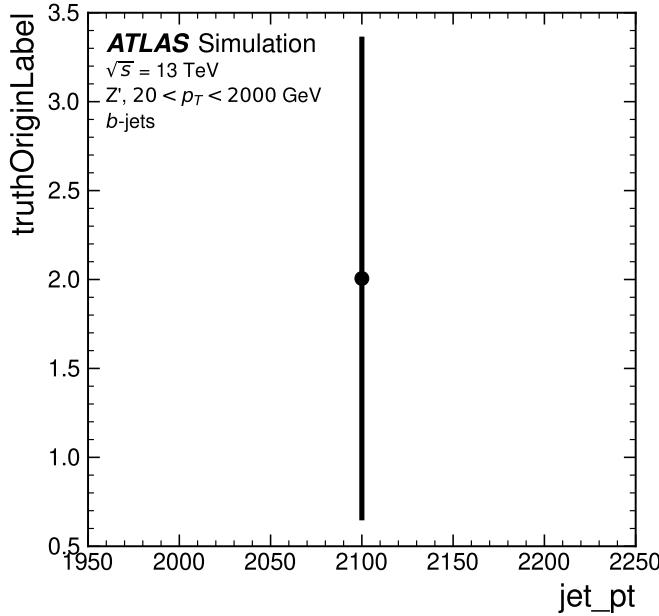


Figure 4.1: The truth b -hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed b -jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.

829 b -hadrons decay weakly to on average four or five collimated stable particles [83].
 830 These particles, along with any other fragmentation particles, are reconstructed in
 831 the detector as a jet. A b -jet has several characteristic features which differentiate
 832 it from light-jets. These features stem from the significant displacement of the
 833 b -hadron that can occur due to its lifetime. The primary feature is the presence of
 834 a high mass secondary vertex that is significantly displaced from the primary vertex.
 835 Reconstruction of these vertices from tracks with common points of spatial origin is
 836 a common approach used in the identification of b -jets.

837 Additional signatures of b -hadrons are as follows. Associated tracks and SVs can have
 838 a large transverse impact parameter d_0 as a result of the b -hadron displacement (as
 839 shown in Fig. 4.2). Since it is common for the b -hadron to decay to a c -hadron with
 840 non-negligible lifetime, tertiary vertices can be found within b -jets resulting from
 841 $b \rightarrow c$ decay chains. The b -hadron also decays semileptonically in approximately 23%
 842 of cases [15]. The presence of a reconstructed electron or muon inside a jet can also
 843 be a key indicator that the jet was instantiated by a b -hadron.

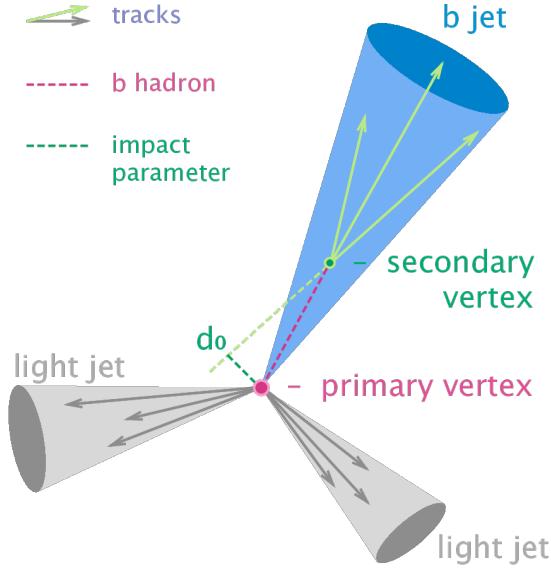


Figure 4.2: Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [84]. The b -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter d_0 is a characteristic property of the trajectories of b -hadron decay products.

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

848 4.1.2 Challenges Facing b -hadron Reconstruction

849 As discussed, a necessary requirement for successful b -tagging is the efficient and
 850 accurate reconstruction of the charged particle trajectories in the jet. For high p_T jets
 851 ($p_T > 200$ GeV) this task becomes difficult due to a combination of effects. As the
 852 b -hadron energy increases, the multiplicity of the fragmentation products inside the
 853 jet increases, while the multiplicity of the products of the weak decay is unaffected.
 854 The “signal” tracks (those from the weak decay of the b -hadron) therefore become
 855 outnumbered. Both fragmentation and b -hadron weak decay products also become
 856 increasingly collimated as their inherited transverse momentum increases. At high
 857 energies, the increased decay length of b -hadrons (and c -hadrons) means that decay
 858 products have less of an opportunity to diverge before reaching the first tracking

859 layers of the detector (shown in Fig. 4.3). If the weak decay of the b -hadron takes
 860 place close enough to a detector layer, or if the particles are otherwise sufficiently
 861 collimated, charge deposits left by nearby particles may not be resolved individually,
 862 instead being reconstructed as merged clusters.

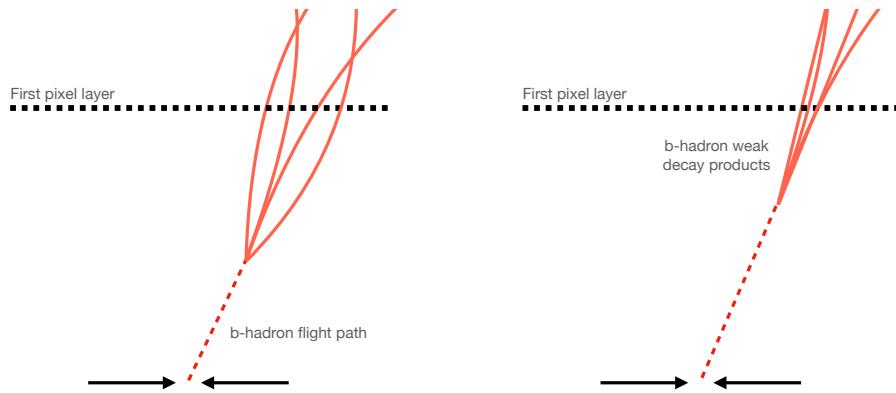


Figure 4.3: At lower p_T (left) the decay length of the b -hadron is reduced, and the resulting decay tracks are less collimated. At higher p_T (right) the b -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching detector elements. As a result, the ID may be unable to resolve charged depositions from different particles, resulting merged clusters.

863 As discussed in Section 3.4.1, merged clusters are generally rare, and so shared
 864 hits generally predict bad tracks and are correspondingly penalised during track
 865 reconstruction. However, in the core of high p_T b -jets the density of particles is high
 866 enough that the probability of cluster merging increases dramatically. Successful
 867 reconstruction of such tracks requires the presence of shared hits to be effectively
 868 dealt with but in the standard reconstruction the the presence of these can end up
 869 impairing the successfully reconstruction of the track. Furthermore, decays may also
 870 take place inside the tracking detectors themselves, which at best leads to missing
 871 measurements on the most sensitive detector layers, and at worst can lead to wrong
 872 inner layer hits being added to displaced tracks, since the reconstruction process
 873 penalises tracks without inner layer hits.

874 The above effects create two related, but distinct problems for b -tagging. The first
 875 part is a drop in track reconstruction efficiency. The presence of shared and missing
 876 hits reduces a track's score in the ambiguity solver meaning that higher ranking, but
 877 potentially worse, 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. 4.4.

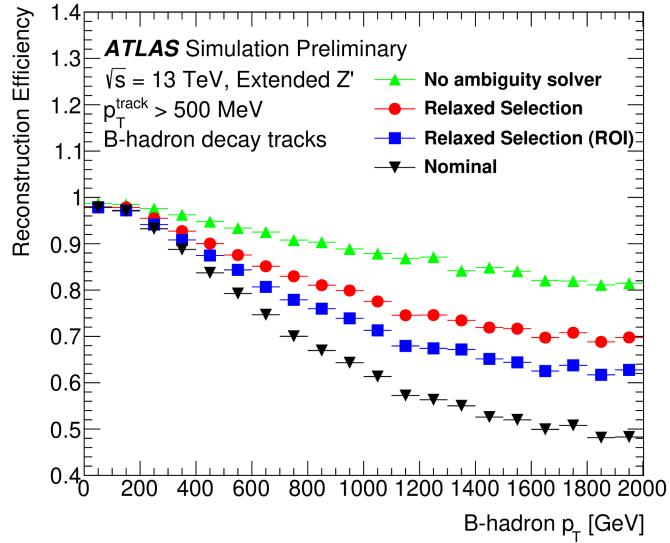


Figure 4.4: b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [85]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- p_T b -hadrons, 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 energy 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. 4.5 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 pseudotrack represent the ideal tracking setup as outlined in Section 3.4.1. The incorrect hits may skew the parameters of the track, which can in turn mislead the

downstream b -tagging algorithms. In particular, b -tagging algorithms rely heavily on the transverse impact parameter significance $s(d_0)$ of the track. 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 downstream b -tagging algorithms.

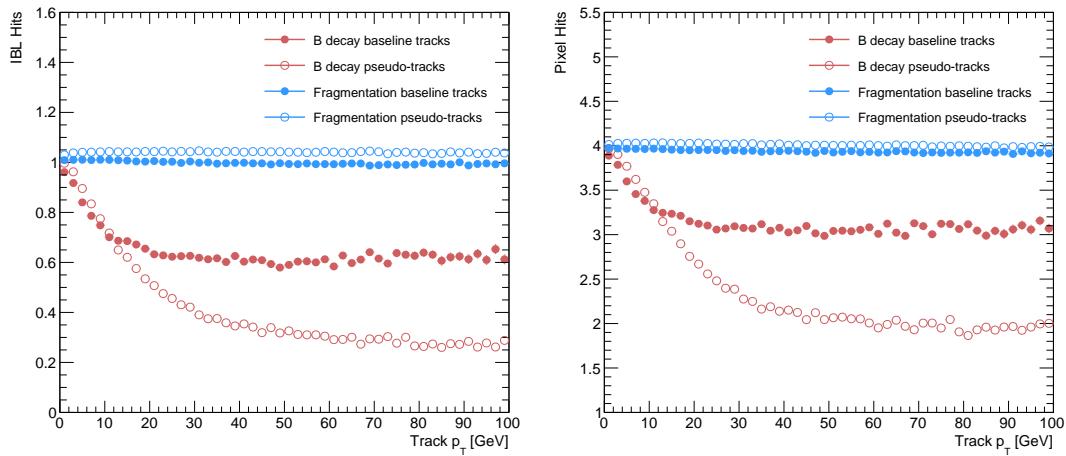


Figure 4.5: Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. Baseline tracks are those produced in the standard 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. 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.

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.

4.2 Investigations into High p_T b -hadron Tracking

In Section 4.2.1 pseudotracks, a key tool for studying the ideal tracking performance of the ATLAS detector, are used to study the shared hit requirements on tracks in

include plot from sebs study showing they are approx similar impacts? or just mention result? Can do put need to remove ATLAS labels. Alternatively you can put an internal reference to his work and state what the outcome is

908 the dense cores of high- p_T b -jets. Section 4.2.2 details a study which investigated
909 modifying the global track fitter to improve reconstruction performance in this
910 regime.

911 4.2.1 Shared Hits

912 The ambiguity solver is not run for pseudotraccks. However, if the standard track
913 collection is produced alongside the pseudotraccks, then cluster splitting neural
914 networks will be run for the standard tracks, and the resulting classification of
915 clusters will be propagated to hits on pseudotraccks. This quirk allows one to study
916 the inefficiencies of the cluster splitting process, and relatedly to determine whether
917 shared hit cuts in the ambiguity solver are too loose or too tight. The fraction of hits
918 that are shared for the IBL and the B-layer is shown in Fig. 4.6. The shared hits on
919 pseudotraccks represent correctly assigned hits from merged clusters that were not
920 able to be classified as split by the cluster splitting neural networks. As such, these
921 represent the number of shared hits the ambiguity solver should aim to allow given
922 the current performance of the cluster splitting algorithm. For shared hits on the IBL
923 for particles produced before the IBL, the baseline selection appears to be successful
924 in disallowing excessive numbers of shared hits. However, the ambiguity solver fails to
925 limit shared hits for those particles produced after the IBL, reflecting the previously
926 discussed problem of displaced tracks picking up incorrect hits. Meanwhile, it is clear
927 that for the B-layer, the ambiguity solver is being overly aggressive in its rejection of
928 shared hits.

929 4.2.2 Global χ^2 Fitter Outlier Removal

930 This section documents ongoing progress into improvement of hit-to-track assignment
931 by using the Global χ^2 Fitter (GX2F) to identify and prevent incorrect hits from
932 being assigned to tracks during the track fit. This is in contrast to a previously
933 investigated approach [86] which attempted to identify and remove wrong hits after
934 the reconstruction of the track. As part of the track fit, an outlier removal procedure
935 is run, in which suspicious hits are identified and removed.

936 The GX2F code, as a relatively low-level component of track reconstruction, has
937 not undergone significant modification for several years, and was originally only

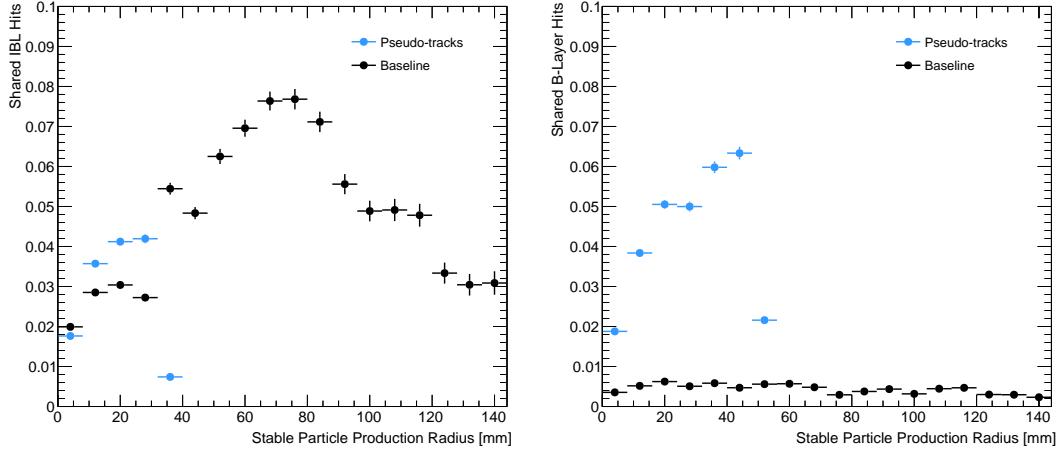


Figure 4.6: The fraction of hits which are shared on b -hadron decay tracks on the IBL (left), and the B-layer (right), as a function of the production radius of the b -hadron decay product. Pseudotrack represent the ideal performance given the ATLAS detector, see Section 3.4.1.

938 optimised in the context of prompt, isolated tracks. During this time, a new tracking
 939 sub-detector, the IBL, was installed. The motivation for looking at the GX2F is that
 940 these changes may require re-optimisation of the GX2F code, and in particular the
 941 outlier removal procedures. Further motivation for this approach comes from the low
 942 rate of labelled outliers in baseline tracking. For example, while approximately 15%
 943 of b -hadron decay tracks have a wrong IBL hit (a value which only increases with
 944 the p_T of the b -hadron), less than 1% of this tracks have had their IBL hit labelled
 945 and removed as an outlier.

946 Implementation

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

955 measurement is calculated according to

$$p_i = \frac{m_i - x_i}{\sqrt{\sigma(m_i)^2 - \sigma(x_i)^2}} \quad (4.2)$$

956 This pull is computed for the transverse and longitudinal coordinates of the mea-
 957 surement, and the maximum of the two is selected and checked to see if it exceeds
 958 a certain selection threshold. If it does, the hit will be removed if the track also
 959 exceeds a threshold on the total χ^2/n . The results of varying the outlier selection
 960 and χ^2/n thresholds are described below.

961 Cut Optimisation

962 A systematic variation of the outlier selection and χ^2/n thresholds has been carried
 963 out. Both thresholds were reduced in fixed step sizes of 0.25 for the outlier selection
 964 threshold and 1 for the χ^2/n threshold. The results for the best performing selections
 965 are discussed below. The value of the outlier selection threshold was reduced from 4
 966 down to 1.75, a change which affects all silicon layers (the TRT has separate outlier
 967 removal logic). Furthermore, a specific cut for the IBL was introduced, and is set
 968 to 1.25. The second threshold on the track χ^2/n was also reduced from 7 to 4.
 969 Finally, instead of taking the maximum of the pulls in the longitudinal and transverse
 970 directions, a quadrature sum is taken of these two values and used. This variation is
 971 labelled “Mod GX2F” in plots.

972 The results are shown in Fig. 4.7 and demonstrate a reduction in wrong hit assignment
 973 whilst also improving slightly the rate at which good hits are assigned to tracks. For a
 974 1 TeV track, the rate to assign good hits to the track increases by approximately
 975 10%, while the rate to assign incorrect hits decreases by approximately 16%. The
 976 improvements are also observed when looking inclusively in all tracks, which avoids
 977 the need for a specific b -jet region-of-interest selection.

978 An improvement, though modest, of all track parameter resolutions and pulls is
 979 observed. The improvement for the transverse impact parameter pull is shown in
 980 Fig. 4.8. The results demonstrate an improvement in hit assignment, unchanged
 981 reconstruction efficiency, and modest improvement in track parameter resolutions
 982 and pulls. In addition, the truth match probability of track is unchanged, suggesting

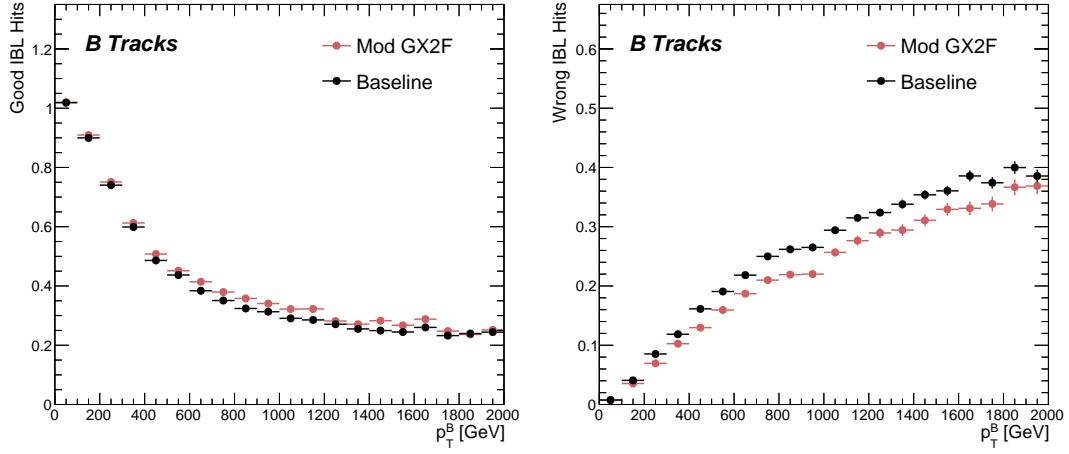


Figure 4.7: The rate to assign good (left) and wrong (right) IBL hits as a function of b -hadron p_T for tracks using baseline tracking (black) and the modified version of the outlier removal procedure (red). For each track, the corresponding p_T bin is filled with the number of good or wrong hits and this value is averaged to show the overall rate.

983 that there is no increase in fake track rates. The changes are expected to have a
 984 negligible impact on computational resources.

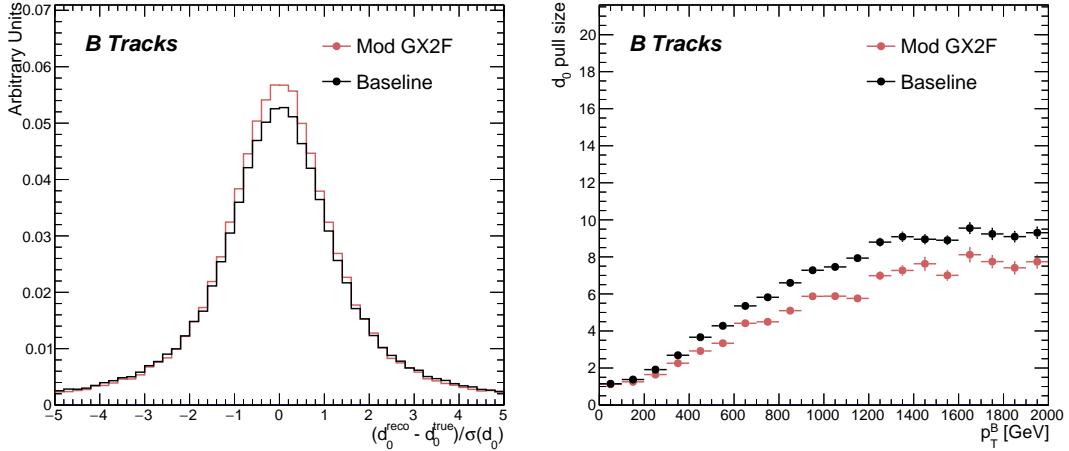


Figure 4.8: (left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified GX2F tracks. (right) The absolute value of the d_0 pull as a function of b -hadron transverse momentum.

985 4.3 Conclusion

986 In this section, the difficulties facing efficient and accurate track reconstruction,
987 and hence performant b -tagging, have been outlined. The ambiguity solver, which
988 attempts to clean or reject tracks which have an excessive number of shared hits,
989 is shown to be overly aggressive in the removal of b -hadron decay product track
990 candidates. The ambiguity solving process relies on a complicated pre-defined
991 selection which has not been optimised for high transverse momentum b -hadron
992 track reconstruction. These conclusions have motivated further ongoing studies into
993 the improvement of the track reconstruction in dense environments and the high- p_T
994 regime, such as those in Ref. [85].

995 An optimisation of the outlier removal process in the global χ^2 fitter was carried
996 out. Though the results show some improvement over the baseline tracking scenario,
997 these results need to be expanded upon by looking at the impact on the downstream
998 b -tagging algorithms before putting them into production. As there are some known
999 data-MC discrepancies, fine tuned optimisation such as the work presented here
1000 presents an opportunity to over-optimise the tracking algorithms on MC. The studies
1001 were carried out in Release 21 of the ATLAS software, and need to be reproduced
1002 using the newer Release 22 to confirm the results against other changes in the baseline
1003 tracking configuration. Thanks to the all-in-one flavour tagging approach described
1004 in Chapter 6, it will also be easier in future to verify that the improvements to the
1005 track reconstruction have a positive impact on the flavour tagging performance.

1006 **Chapter 5**

1007 **Track Classification MVA**

1008 The chapter details work on implementing a multivariate algorithm (MVA) to predict
1009 the truth origin of reconstructed tracks. An introduction to formalisms of machine
1010 learning is given in Section 5.1. In Section 5.2, the truth origin label is defined,
1011 and in Section 5.3 these labels are used to train a machine learning model that can
1012 effectively discriminate between good and fake tracks. Several studies motivated this
1013 work by demonstrating that at high p_T , b -tagging performance was degraded by the
1014 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm
1015 could be trained to detect fake tracks, these could be removed before their input to
1016 the b -tagging algorithms with the aim of improving performance.

1017 **5.1 Machine Learning Background**

1018 Over the past few decades, machine learning (ML) techniques have become increasing-
1019 ly popular in high energy physics experiments due the increased volumes of
1020 high-dimensional data and improvements in the techniques used (in particular deep
1021 learning). Machine learning is the process by which a computer program uses data
1022 to learn suitable parameters for a predictive model. This is opposed to explicitly
1023 providing instructions on how to perform a task. A subfield known as *supervised*
1024 *learning* is used in this work, and consists of exposing a model to a large number of
1025 labelled examples in order to extract relationships between the input data and their
1026 labels. These relationships are often complex, and explicitly programmed rules can
1027 fail to fully capture the relationships between inputs and outputs.

1028 In the simplest case, a set of m labelled training examples $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$
 1029 is collected. Each element (x_i, y_i) consists of a input vector $x_i \in \mathbb{R}^{\text{input}}$, and the
 1030 corresponding label y_i . In classification problems, these labels are integer *class*
 1031 *labels* $y_i \in \{0, \dots, N - 1\}$, where N is the number of classes, which specify which
 1032 of a pre-determined set of categorical classes the training example belongs to. The
 1033 rest of the discussion in this chapter is limited to binary classification problems
 1034 ($N = 2$). The two classes are often referred to as signal ($y_i = 1$) and background
 1035 ($y_i = 0$), which need to be separated. Collecting sufficient and suitable data is one
 1036 of the primary challenges of machine learning, as such data is not always readily
 1037 available. Fortunately, sophisticated tools to simulate particle collisions have already
 1038 been developed by the scientific community [87, 88]. These tools play a key role in
 1039 generating a suitably large amount of labelled data which is used to train algorithms.
 1040 More detail on the input datasets is given in Section 5.3.1.

1041 After obtaining suitable training data, the next step is to define a model. Given an
 1042 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
 1043 parameterised functional mapping from input space to output space. Given an input
 1044 example x_i and a set of parameters θ , the model outputs a prediction $\hat{y}_i \in (0, 1)$ for
 1045 the true label y_i , as in

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

1046 The output \hat{y}_i is in the interval $(0, 1)$ so as to be interpreted as the probability that
 1047 the input example x_i belongs to the signal class. The parameters θ of the model are
 1048 randomly initialised, and the model is designed to be expressive enough to correctly
 1049 map the inputs x_i to the outputs y_i given a reasonable optimisation of the parameters.
 1050 To perform this optimisation, the model is then trained, which amounts to showing
 1051 the model a series of labelled training examples and modifying the parameters of the
 1052 model based on its ability to correctly predict the labels.

1053 5.1.1 Neural Networks

1054 Neural networks (NNs) are a common choice for the machine learning model f since
 1055 they have the ability to approximate any function [89] and are easy to train via
 1056 backpropagation [90].

1057 **Artificial Neurons**

1058 The basic functional component of a NN is the *artificial neuron* or node, which is
1059 loosely inspired by a mathematical model of a biological neuron [91, 92]. A diagram
1060 of an artificial neuron is shown in Fig. 5.1 Each neuron is defined by its parameters
1061 or *weights* θ and a choice of activation function. Each neuron takes a fixed number
1062 of inputs and computes the dot product of the input and weight vectors $x^T \theta$ and
1063 additionally adds a constant bias term θ_0 . This term plays the role of a trainable
1064 constant value that is independent of the inputs.

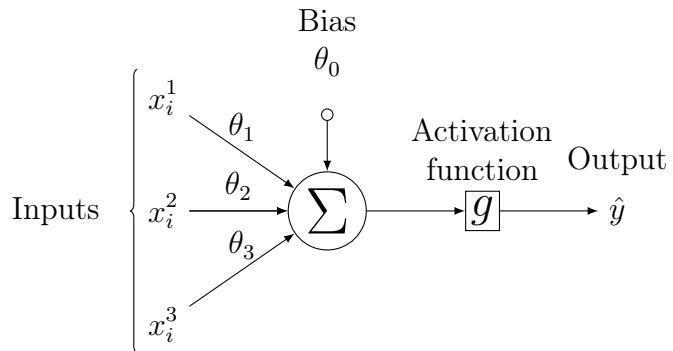


Figure 5.1: A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result z is passed to an activation function. Each neuron can be thought of as a logistic regression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

1065 The output of the dot product and bias term z is fed into an activation function
1066 g . The activation function has several uses, most notably acting as a source of
1067 non-linearity and bounding the output of the neuron. Some common activation
1068 functions (sigmod, tanh, ReLU and SiLU) are shown in Fig. 5.2. The choice of
1069 activation function can have implications for the performance and convergence of
1070 the network, since the gradient of g is used to compute the weight updates during
1071 training. This is also why input data is typically normalised to have zero mean and
1072 unity variance [93].

1073 **Networks**

1074 Several neurons are linked together in layers to form a neural network. The inputs
1075 are propagated layer-by-layer through the network until reaching the final output

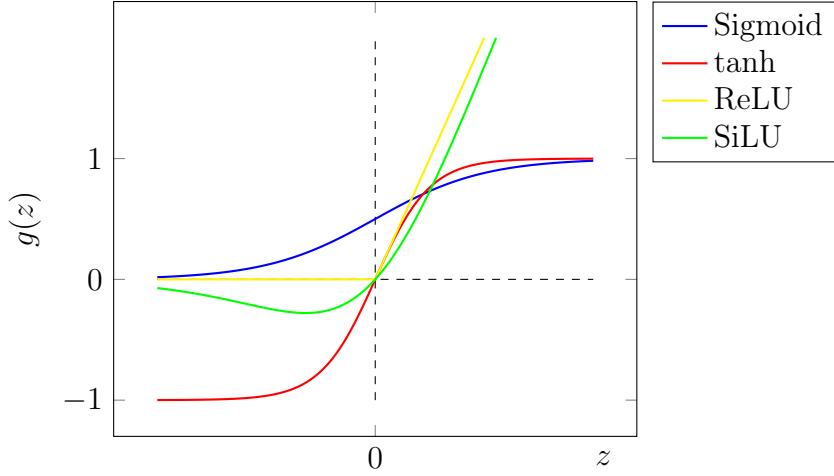


Figure 5.2: The output of several common choices for the activation function $g(z)$ of an artificial neuron. The input z is the output of the dot product between the activation and the weights, plus a bias term.

layer. The number of layers and neurons per layer are important hyperparameters (those parameters which are not optimised as part of the training process) which influence the performance of the model. In the case of binary classification, the final output layer consists of a single neuron with a sigmoid activation

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

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

5.1.2 Training with Gradient Descent

A training algorithm is used to optimise the weights 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

1091 problems, the binary cross entropy given in Eq. (5.3) loss if often used.

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

1092 Since the model f is differentiable, the error for each parameter θ_i can be computed by
1093 taking the partial derivative of L with respect to the parameter. Updated parameters
1094 θ'_i are calculated by updating the original parameter in the direction which reduces
1095 the loss.

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

1096 The hyperparameter α is known as the *learning rate* and dictates the size of the
1097 step taken in the direction of the slope. The errors for each parameter are efficiently
1098 calculated using the backpropagation algorithm [90]. The process of updating weights
1099 is repeated until the weights converge, which means the network is trained. In practice,
1100 small batches of the input data are shown to the network at a time. For each batch
1101 the average loss is calculated and the network's weights are updated. There are many
1102 extensions and variations of the gradient descent algorithm. This work uses the Adam
1103 optimiser which adds momentum to the weight updates (dampening oscillations)
1104 and an adaptive per-parameter learning rate [94].

1105 5.2 Track Truth Origin Labelling

1106 Crucial to supervised learning techniques are the ground truth class labels which the
1107 machine learning model is trained to predict. A set of track truth labels which a
1108 high degree of granularity have been implemented in the ATLAS software stack, and
1109 are listed in Table 5.1. The labelling scheme has been designed to be useful beyond
1110 the classification of good and fake tracks. The origins are determined by analysing
1111 the simulated record to determine the physical process that led to the creation of
1112 the truth (i.e. simulated) particle which is associated with each reconstructed track.
1113 Tracks are associated with truth particles by selecting the particle with the highest
1114 *truth-matching probability* (TMP), defined in Eq. (5.5). This is a weighted sum of
1115 the number of hits on a reconstructed track which are from the same truth particle,
1116 versus the total number of hits on the track. The weights are subdetector-dependent
1117 and are designed to account for the varying importance of the different subdetectors

1118 (based upon their precision) in the reconstruction of a track.

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

1119 For the fake track classification tool, the track truth origins in Table 5.1 are used
1120 to construct a binary label by assigning all fake tracks to the background category,
1121 and all other tracks as signal. The fake track classifier is then trained to distinguish
1122 between these two categories of tracks. Fake tracks are defined using the TMP,
1123 with a $\text{TMP} < 0.75$ ¹ giving a track the label of fake. Fake tracks are made up of
1124 combinatorial fakes, which are tracks which do not correspond to the trajectory of
1125 any truth particle, and poorly reconstructed tracks, which may somewhat resemble
1126 the trajectory of a truth particle due to the presence of some wrong hits on the track,
1127 will not accurately reproduce a true trajectory.

Truth Origin	Description
Pileup	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 physics process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [56]. A truth-matching probability of less than 0.5 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.

1128 5.3 Fake Track Identification Tool

1129 The rate of fake tracks increases at high transverse momentum as shown in Fig. 5.3 due
1130 to the difficulties in track reconstruction outlined in Section 4.1.2. The performance

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

of b -tagging algorithms is reduced as a direct result of the presence of these tracks as shown for SV1 (see Section 3.4.2) in Fig. 5.4, where the light-jet efficiency decreases by up to 35% at a b -jet efficiency of 35%.

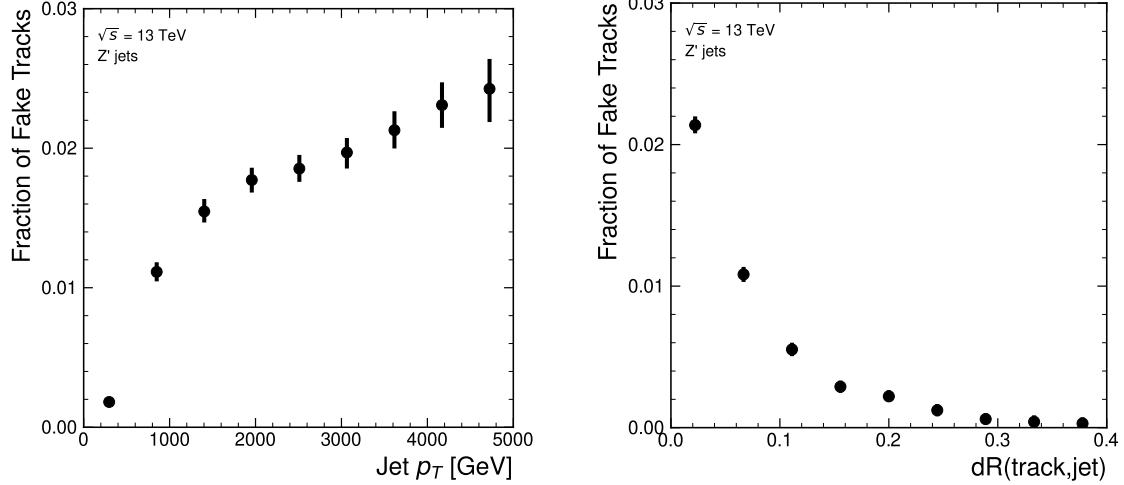


Figure 5.3: Rate of fake tracks as a function of jet transverse momentum (left) and ΔR (track, jet) (right). The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases.

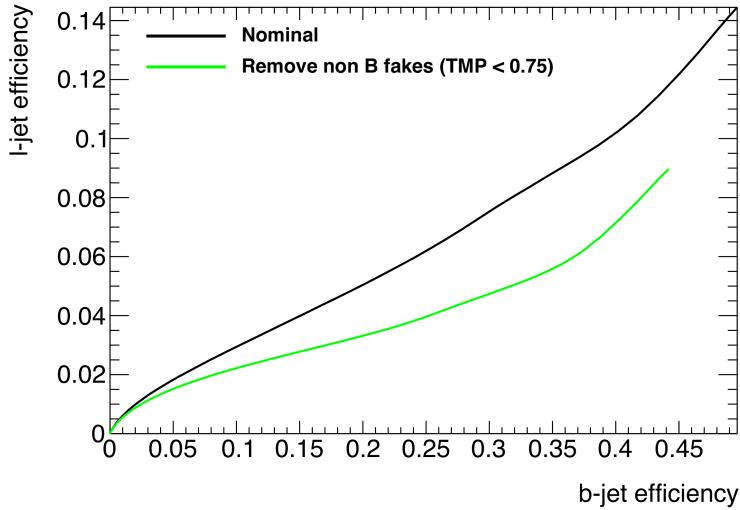


Figure 5.4: The light-jet efficiency of the low level tagger SV1 for jets in the Z' sample with $250 < p_T < 5000$ GeV, as a function of b -jet efficiency. The nominal tracking setup (black) is shown alongside the case where fake tracks which are not from the decay of a b -hadron are removed. The light-jet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

1134 To identify and remove fake tracks, a NN classification tool was trained with all
1135 non-fake tracks as the signal class and fake tracks as the background class. Inputs to
1136 the model are described in Section 5.3.2, while fake track removal performance is
1137 given in Section 5.3.4.

1138 **5.3.1 Datasets**

1139 To train and evaluate the model, simulated SM $t\bar{t}$ and BSM Z' events initiated by
1140 proton-proton collisions at a center of mass energy $\sqrt{s} = 13$ TeV are used. The Z'
1141 sample is constructed in such a manner that it has a relatively flat jet p_T spectrum
1142 up to 5 TeV and decays democratically to equal numbers of b -, c - and light-jets.
1143 The generation of the simulated event samples includes the effect of multiple pp
1144 interactions per bunch crossing with an average pileup of $\langle \mu \rangle = 40$, which includes
1145 the effect on the detector response due to interactions from bunch crossings before
1146 or after the one containing the hard interaction.

1147 The $t\bar{t}$ events are generated using the POWHEGBOX [95–98] v2 generator at next-to-
1148 leading order with the NNPDF3.0NLO [99] set of parton distribution functions
1149 (PDFs). The h_{damp} parameter² is set to 1.5 times the mass of the top-quark
1150 (m_{top}) [100], with $m_{\text{top}} = 172.5$ GeV. The events are interfaced to PYTHIA 8.230 [101]
1151 to model the parton shower, hadronisation, and underlying event, with parameters
1152 set according to the A14 tune [102] and using the NNPDF2.3LO set of PDFs [103].
1153 Z' events are generated with PYTHIA 8.2.12 with the same tune and PDF set. The
1154 decays of b - and c -hadrons are performed by EVTGEN v1.6.0 [104]. Particles are
1155 passed through the ATLAS detector simulation [105] based on GEANT4 [106].

1156 Jets are required to have a pseudorapidity $|\eta| < 2.5$ and $p_T > 20$ GeV. Additionally, a
1157 standard selection using the Jet Vertex Tagger (JVT) algorithm at the tight working
1158 point is applied to jets with $p_T < 60$ GeV and $|\eta| < 2.4$ in order to suppress pile-up
1159 contamination [107].

²The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and thus effectively regulates the high- p_T radiation against which the $t\bar{t}$ system recoils.

1160 5.3.2 Model Inputs

1161 The fake track MVA is given two jet variables and 20 tracking related variables
1162 for each track fed into the network. The jet transverse momentum and signed
1163 pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in
1164 Table 5.2. The track parameters and hit pattern are key indicators of whether or
1165 not a track is fake. The FracRank variable is the ordered index of the tracks that
1166 pass the ambiguity solver’s selection divided by the total number of successfully
1167 reconstructed tracks in the event. The ambiguity solver processes track candidates
1168 iteratively in order of an internal score (see Section 3.4.1), and the order in which
1169 tracks are accepted is preserved. Since tracks with shared hits have lower scores,
1170 tracks which do not require the removal of shared hits are likely to be processed
1171 and accepted earlier on, whereas tracks with shared hits will be processed later and
1172 potentially have their shared hits removed. Hence the FracRank variable gives an
1173 indication of the of how easy it was for the track to be reconstructed.

1174 Track selection follows the loose selection described in Ref. [80] and outlined in
1175 Table 5.3, which was found to improve the performance compared to previous tighter
1176 selections, whilst ensuring good resolution of tracks and a low fake rate [56]. Inputs
1177 are scaled to have a central value of zero and a variance of unity before training and
1178 evaluation.

1179 5.3.3 Model Hyperparameters

1180 Due to the imbalance between the two classes (with fake tracks being relatively
1181 uncommon), a weight was added to the loss function for the background class to
1182 account for this. The NN was made up of two hidden layers with 220 nodes per layer.
1183 The ReLU activation function was used in conjunction with the Adam optimiser with
1184 a learning rate of $1e-3$. Optimisation of the networks architecture was carried out to
1185 ensure optimal performance with a relatively small number of learnable parameters –
1186 54 thousand. The model was trained using 40 million tracks with a futher 1 million
1187 tracks each used for validation and testing. A full list of the model hyperparameters
1188 is given in Table 5.4.

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 [56], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.

Parameter	Selection
p_T	> 500 MeV
$ d_0 $	< 3.5 mm
$ z_0 \sin \theta $	< 5 mm
Silicon hits	≥ 8
Shared silicon hits	< 2
Silicon holes	< 3
Pixel holes	< 2

Table 5.3: 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). Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [56]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

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 5.4: Hyperparameter for the track classification model

5.3.4 Results

In order to evaluate the fake track classification tool, a orthogonal test sample of 1 million tracks in jets in the combined $t\bar{t}$ and Z' samples is 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.

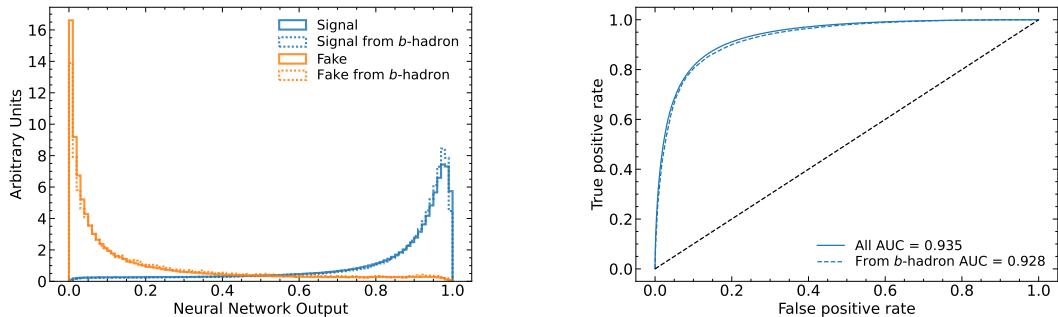


Figure 5.5: (left) Normalised histogram of the model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed line). The model is able 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 5.5. 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 5.5 also shows that a significant amount of tracks which are labelled as both fake and from the decay of a b -hadron are also removed. This can happen because fake tracks with $TMP < 0.75$ are still matched to a truth particle, 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 5.5: 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. The continuous output of the model allows for the tuning of good and fake track identification efficiencies.

5.4 b -hadron Track Identification

After initial tests and investigation, it was found that fake tracks which were the result of b -hadron decays actually aided b -tagging performance. The application of a single tool which removed all fake tracks was therefore not optimal. A second tool was therefore trained in the same manner as the first, this one was designed to distinguish between those tracks which were from the decay of a b -hadron (FromB and FromBC in Table 5.1) and those which were not (all other truth origins). The b -hadron decay track MVA was trained using the same setup as described above, with the same tracks, input variables, and training procedure. The performance of the model to separate b -hadron decay tracks from other tracks is shown in Fig. 5.6. Using a selection WP of 0.1, the model can retain 98.5% of b -hadron tracks and reject 46.2% of tracks not from the decay of a b -hadron. In Section 5.5, this model is used in conjunction with the fake track identification MVA described in Section 5.3.4 to identify and remove fake tracks which are not from the decay of a b -hadron.

5.5 Combined Approach

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

The light-jet efficiency of SV1 is successfully reduced when using the combined tools to remove fake tracks that are not from a b -hadron decay, as shown in Fig. 5.7. At a b -jet efficiency of 70%, the light-jet mistag rate for jets with $250 < p_T < 400 \text{ GeV}$

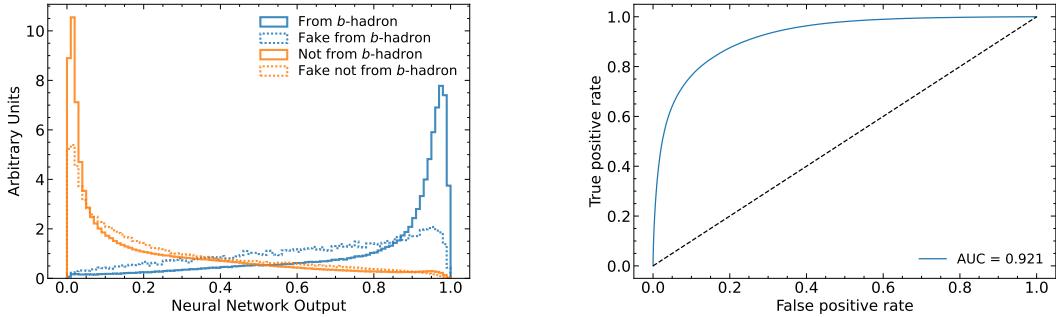


Figure 5.6: (left) Normalised histogram of the model output separated for tracks from the decay of a b -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line).

WP	Fake Cut	MVA	b -hadron Decay MVA Cut	Retained Tracks	b -hadron	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 5.6: 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).

is reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For jets with $400 < p_T < 1000 \text{ GeV}$ the mistage rate drops from 0.1 to 0.08 for a similar relative improvement of 20%. The performance of the fake track removal approach was also tested for the other low level vertexing algorithm: JetFitter. A similar level of improvement in the light-jet mistag rate was observed of up to a 20% reduction for both low- and high- p_T jets in the Z' sample. Together, these results demonstrate that by identifying and removing fake tracks which are not the result of the weak decay of a b -hadron, the performance of the low level tagging algorithms can be improved.

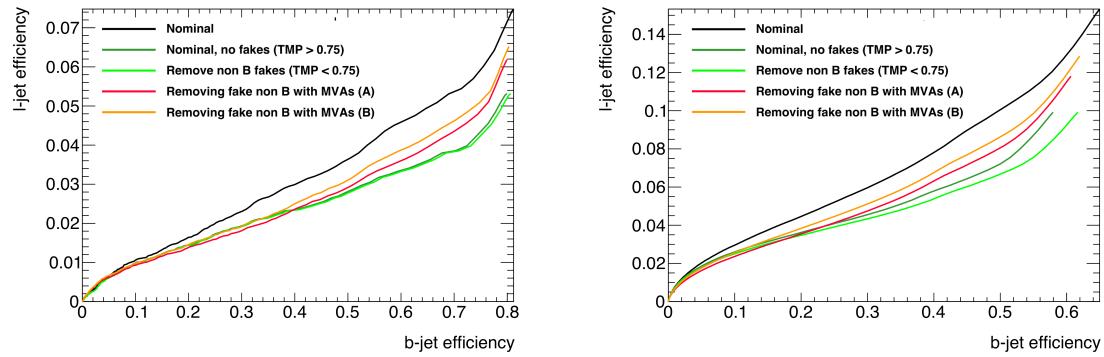


Figure 5.7: The effect of applying the fake track identification algorithm together with the b -hadron decay track identification on the jet tagging performance of SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). 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. 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

1246 mistagging rate of SV1 and JetFitter by up to 20% at high transverse momentum.
1247 The improvement achieved using the classification tools was in general comparable with
1248 the improvement achieved when using the truth information to remove the fake
1249 tracks not from the decay of a b -hadron.

1250 **Future Work**

1251 While removing tracks prior to their input to the low level tagging algorithms is
1252 shown here to be beneficial, a more performant alternative might be to keep these
1253 tracks but label them as being fake (for example using the output of the classification
1254 tool), and allow the tagging algorithms to take this into consideration, potentially
1255 making use of this information. This is not straightforward with manually optimised
1256 taggers such as SV1 and JetFitter, but is possible with more advanced taggers as
1257 described in Chapter 6.

1258 Tools which identify the origin of a given track have other potential uses. One
1259 application is to isolate a relatively pure sample of fake tracks which can be used
1260 to estimate the fake track rate in data, which would be useful for estimating the
1261 uncertainty on fake track modelling. Another application would be to use the
1262 b -hadron track identification tool to improve the track-to-jet association. Both
1263 applications are currently under investigation.

1264 The approach here works on a track-by-track basis, but a more sophisticated approach
1265 would consider the correlations between the tracks inside a jet, as shown in Chapter 6.

1266 Also left for future work is to simultaneously train a single tool which discriminates
1267 between all the truth origins listed in Table 5.1. Such a tool would be useful as a
1268 general purpose multiclass classifier.

₁₂₆₉ **Chapter 6**

₁₂₇₀ **Graph Neural Network Flavour
Tagger**

₁₂₇₂ As discussed in Chapter 4, flavour tagging is the identification of jets instantiated
₁₂₇₃ from b - and c -hadrons. Flavour tagging is a critical component of the physics
₁₂₇₄ programme of the ATLAS experiment. It is of crucial importance for the study of the
₁₂₇₅ Standard Model (SM) Higgs boson and the top quark, which decay preferentially to
₁₂₇₆ b -quarks [108, 109], and additionally for several Beyond the Standard Model (BSM)
₁₂₇₇ resonances that readily decay to heavy flavour quarks [110].

₁₂₇₈ Existing flavour tagging algorithms, such as DL1r [77, 111], use as inputs the outputs
₁₂₇₉ from a number of low-level algorithms. The low-level algorithms each reconstruct
₁₂₈₀ various features using the tracks which have been associated to the jet. The outputs
₁₂₈₁ from the low-level algorithms are then combined in a machine learning model. This
₁₂₈₂ chapter introduces GN1, a novel ML-based flavour tagging algorithm based on graph
₁₂₈₃ neural networks (GNNs).

₁₂₈₄ As opposed to the previous two-tiered approach described in Chapter 4, which requires
₁₂₈₅ the use of both low- and high-level algorithms, GN1 takes as inputs information
₁₂₈₆ from an unordered variable number of tracks as input, to predict the jet flavour
₁₂₈₇ without requiring outputs from the intermediate low-level algorithms. In addition
₁₂₈₈ to predicting the flavour of the jet, the model predicts which physical processes
₁₂₈₉ produced the various tracks, and groups the tracks into vertices. These auxiliary
₁₂₉₀ training objectives provide valuable additional information about the contents of the
₁₂₉₁ jet and enhance the performance of the primary flavour prediction task.

1292 In Section 6.1, an overview of the approach used for GN1 is provided. An introduction
 1293 to the theory of GNNs is provided in Section 6.2. Details of the experimental setup
 1294 are provided in Section 6.3, while the architecture of GN1 is specified in Section 6.4.3.
 1295 In Section 6.4.4, the training procedure is described, and in Section 6.5 the results
 1296 are shown.

1297 6.1 Motivation

1298 GN1 is a monolithic approach to flavour tagging as illustrated in Fig. 6.1. The use
 1299 of GNNs offers a natural way to classify jets with variable numbers of unordered
 1300 associated tracks (see Section 6.2), while allowing for the inclusion of auxiliary
 1301 training objectives [112, 113].

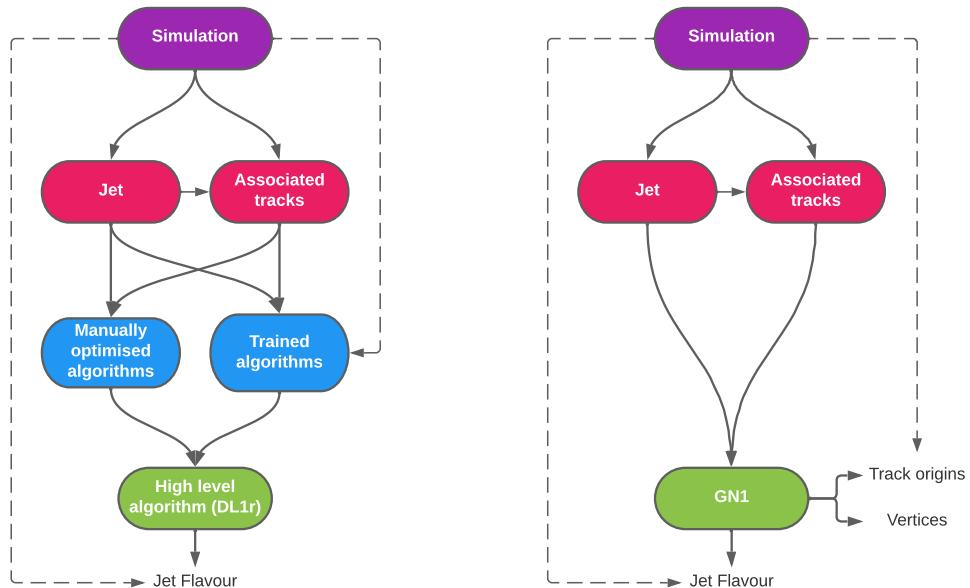


Figure 6.1: Comparison of the existing flavour tagging scheme (left) and GN1 (right) [2]. The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, 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.

1302 As described in Chapter 4, current flavour tagging algorithms utilise a two-tired
1303 approach. The high-level tagger DL1r outputs variables which provides good discrim-
1304 ination between the different jet flavours. In contrast GN1 consists of only a single
1305 neural network, which the tracks as inputs along with some kinematic information
1306 about the jet. As a result, it does not depend on the outputs of any other flavour
1307 tagging algorithm. A simple training of the model fully optimises its parameters,
1308 represting a significant simplification with respect to the optimisation procedure for
1309 DL1r. This is particularly important when optimising the tagger for new regions of
1310 phase space (e.g. c -tagging or high- p_T b -tagging), or when the detector is upgraded
1311 or the charged particle reconstruction or selection algorithms are re-optimised.

1312 GN1 is trained to learn about the internal structure of the jet through the use of two
1313 auxiliary training objectives: the prediction of the underlying physics process from
1314 which each track originated, and the grouping of tracks originating from a common
1315 spatial position (i.e. a common vertex). These auxiliary objectives are meant to
1316 guide the neural network towards a more complete understanding of the underlying
1317 physics inside the jet, thereby removing the need for the low-level algorithms, which
1318 previously contained information about the underlying physics in their design. The
1319 training targets for the primary and auxiliary objectives are extracted from truth
1320 information, i.e. information that is only available in simulation, as opposed to
1321 reconstructed quantities available in both collision data and simulation.

1322 In this chapter, the following advantages of the GN1 approach will be demonstrated:

- 1323 1. GN1 boasts improved performance with respect to the current ATLAS flavour
1324 tagging algorithms, with significantly larger background rejection rates for a
1325 given signal efficiency. Alternatively the rejection rates can be kept fixed for a
1326 substantial increase in signal efficiency, in particular at high- p_T .
- 1327 2. The same network architecture can be easily optimised for a wider variety of
1328 use cases (e.g. c -jet tagging and high- p_T jet tagging) since there are no low-level
1329 algorithms to retune.
- 1330 3. There are fewer algorithms to maintain.
- 1331 4. Alongside the network's prediction of the jet flavour, the auxiliary vertex and
1332 track origin predictions provide more information on why a jet was (mis)tagged

or not. This information can also have uses in other applications, for instance to explicitly reconstruct displaced decay vertices or to remove fake tracks.¹

6.2 Graph Neural Network Theory

Graph neural networks are a more sophisticated neural network model (see Section 5.1.1) that are designed to operate on graph structured data. A brief introduction to GNNs is provided in this section following the formalism in Ref. [114].

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

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

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

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

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

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

¹A fake track is defined as a track with a truth-matching probability less than 0.5, where the truth-matching probability is defined in Ref. [56].

1356 using the global network ϕ^u as in

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

1357 where \bar{e}' is the aggregation $\rho^{e \rightarrow u}$ over all updated edges e'_i and \bar{h}' is the aggregation
1358 $\rho^{h \rightarrow u}$ over all updated nodes h'_i .

1359 The graph network layer performs a graph convolution, in an analogous way to a
1360 convolutional neural network operating on a grid of pixels. The above description is
1361 general, and not all concrete implementations of GNNs need implement every aspect.
1362 For example, the global graph representation need not be present, and it is also
1363 possible that no dedicated edge features are present. In such cases the corresponding
1364 update and aggregation functions are not needed. Fig. 6.2 shows two possible graph
1365 network update layers. The layer used in the GN1 model is specified in more detail
1366 in Section 6.4.3.

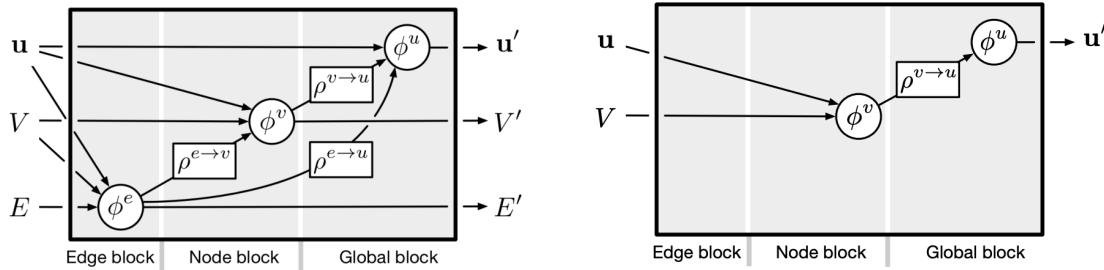


Figure 6.2: The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [115]. Reproduced from Ref. [114].

1367 6.3 Experimental Setup

1368 6.3.1 Datasets

1369 The datasets used to train the GN1 tagger are the same as described in Section 5.3.1.
1370 The training dataset contains 30 million jets, 60% of which are $t\bar{t}$ jets and 40%
1371 of which are Z' jets. In order to evaluate the performance of the model during,
1372 a statistically independent set of 500k testing jets from both the $t\bar{t}$ and Z' samples are
1373 used. For the final testing of the model and the creation of the performance plots,

1374 a further 1 million independent testing jets from each of the $t\bar{t}$ and Z' samples are
1375 used. Before being fed into the model, the track- and jet-level inputs are normalised
1376 to have a mean of zero and a variance of unity. The jet flavour labels are assigned
1377 as described in Section 3.4.3. Truth labelled b -, c - and light-jets are kinematically
1378 re-sampled in p_T and η to ensure identical distributions in these variables.

1379 6.4 Model Architecture

1380 6.4.1 Model Inputs

1381 As inputs, GN1 is fed two kinematic jet variables and an unordered set of up to 40
1382 tracks which have been associated to the jet. Each track consists of 21 variables. The
1383 kinematic jet variables are the jet transverse momentum and signed pseudorapidity.
1384 The input variables which are provided for each track are listed in Table 6.1. For
1385 each track, variables containing the track parameters and uncertainties, and detailed
1386 information on the hit content are provided as inputs to the model.

1387 In cores of high- p_T jets, track density is high due to the increased multiplicity and
1388 collimation of tracks (see Chapter 4). As a result, the separation between tracks
1389 can be of the same order as the active sensor dimensions, resulting in an increase
1390 in merged clusters and tracks which share hits [56]. Due to the relatively long
1391 lifetimes of b -hadrons and c -hadrons, which can traverse several layers of the ID
1392 before decaying and have highly collimated decay products, the presence of shared
1393 or missing hits is a critical signature of heavy flavour jets.

1394 Dependence of the model on the absolute value of the azimuthal jet angle ϕ is
1395 explicitly removed by providing only the azimuthal angle of tracks relative to the jet
1396 axis. The track pseudorapidity is also provided relative to the jet axis.

1397 Since heavy flavour hadrons can decay semileptonically approximately 20% of the time,
1398 the presence of a reconstructed lepton in the jet carries discriminating information
1399 about the jet flavour. To exploit this, a variant of GN1 called GN1Lep is trained in
1400 addition to the baseline model. The GN1Lep variant is identical to the baseline model,
1401 except for the inclusion an additional track-level input, leptonID, which indicates
1402 if the track was used in the reconstruction of an electron, a muon or neither. The

1403 variable is signed by the charge of the reconstructed lepton. The leptons used in the
1404 definition of the leptonID variable are required to satisfy basic quality requirements.
1405 The muons are required to be combined [116], and the electrons are required to pass
1406 the *VeryLoose* likelihood-based identification working point [117].

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 [2]. 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 [56], 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.

¹⁴⁰⁷ The selections applied to the tracks is the same as that used for the fake track
¹⁴⁰⁸ classification MVA described in Chapter 5. The full set of track selections is listed
¹⁴⁰⁹ in Table 5.3. This selection was found to improve the flavour tagging performance
¹⁴¹⁰ compared to previous tighter selections, whilst ensuring good resolution of tracks
¹⁴¹¹ and a low fake rate [56]. However, Section 6.5.8 demonstrates that further relaxation
¹⁴¹² of the track selection requirements may be warranted.

¹⁴¹³ If more than 40 tracks are associated to a given jet, only the first 40 tracks with the
¹⁴¹⁴ largest transverse IP significance² $s(d_0)$ are fed into the model as inputs.

¹⁴¹⁵ 6.4.2 Auxiliary Training Objectives

¹⁴¹⁶ In addition to the jet flavour classification, two auxiliary training objectives are
¹⁴¹⁷ defined. The first auxiliary objective is the prediction of the physical process that
¹⁴¹⁸ gave rise to each track within the jet (i.e. the track origin), while the second is the
¹⁴¹⁹ prediction of track-pair vertex compatibility. Each auxiliary training objective comes
¹⁴²⁰ with a training target which, similar to the jet flavour label, is a truth labels derived
¹⁴²¹ from the simulation. The presence of the auxiliary training objectives improves the
¹⁴²² jet classification performance as demonstrated in Section 6.5.3.

¹⁴²³ For the track origin prediction objective, each track is labelled with one of the
¹⁴²⁴ exclusive categories defined in Table 5.1 of Section 5.2 after analysing the particle
¹⁴²⁵ interaction (or lack thereof) which led to its formation. Since the presence of different
¹⁴²⁶ track origins is strongly related to the flavour of the jet, training GN1 to recognise
¹⁴²⁷ the origin of the tracks provides an additional handle on the classification of the
¹⁴²⁸ jet flavour. This task may also aid the jet flavour prediction by acting as a form of
¹⁴²⁹ supervised attention [118] - in detecting tracks from heavy flavour decays the model
¹⁴³⁰ may learn to pay more attention to these tracks.

¹⁴³¹ The vertexing auxiliary objective makes use of the fact that displaced decays of b -
¹⁴³² and c -hadrons lead to secondary and tertiary vertices inside the jet, as described in
¹⁴³³ Section 4.1.1. The presence of displaced secondary vertices is not a completely clean
¹⁴³⁴ signal of a heavy flavour jet, as displaced secondary vertices can also occur in light-jets
¹⁴³⁵ as a result of material interactions, conversions, and long-lived particle decays (e.g.

²Impact parameter significances are defined as the IP divided by its corresponding uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. Track IP significances are lifetime signed according to the track's direction with respect to the jet axis and the primary vertex [38].

K_S^0 and Λ^0). For the auxiliary object, GN1 predicts a binary label for each pair of tracks in the jet. The label has a value of 1 if the truth particles associated with the two tracks in the pair originated from the same spatial point, and 0 otherwise. To derive the corresponding truth labels for training, truth production vertices within 0.1 mm are merged. Track-pairs where one or both of the tracks in the pair have an origin label of either Pileup or Fake are given a label of 0. Using the pairwise predictions from the model, groups of tracks that have common compatibility can be formed, resulting in the finding of vertices. Two existing low-level tagging algorithms, SV1 and JetFitter (introduced in Section 3.4.2), are currently used to find and reconstruct vertices inside jets and are used as inputs to the existing jet flavour tagger DL1r. The addition of this auxiliary training objective removes the need for inputs from a dedicated secondary vertexing algorithm.

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

6.4.3 Architecture

As discussed in the previous sections, GN1 is a graph neural network which makes use of auxiliary training objectives in order to determine the jet flavour. A coarse optimisation of the network architecture hyperparameters (for example number of layers and number of neurons per layer) has been carried out in order to maximise the flavour tagging performance, but it is likely that further dedicated optimisation studies could lead to further performance improvements.

The model architecture builds on a previous implementation of a GNN-based jet tagger [113]. The previous approach was comprised of two separate graph neural networks with the auxiliary tasks being performed at an intermediate stage after the first and before the second. This two stage approach was found to be unnecessary and as such GN1 simplifies the architecture into a single graph neural network with the auxiliary tasks being performed at the end, alongside the primary jet classification task. GN1 makes use of a more sophisticated graph neural network layer [119],

1467 which is described in more detail below. The changes significantly improved tagging
1468 performance and also led to a significant reduction in training time.

1469 As inputs, the model takes information about the jet and a number of associated tracks,
1470 as detailed in Section 6.4.1. The jet variables are concatenated with the
1471 variables for each track as shown in Fig. 6.3. The combined jet-track input vectors
1472 are then fed into a per-track initialisation network with three hidden layers, each
1473 containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 6.4. The
1474 track initialisation network is similar to a deep sets model [115], but does not include
1475 a reduction operation (mean or summation) over the output track representations.
1476 The initialisation network allows for initial per-track input processing without the
1477 associated parameter count cost of the graph convolutional layers described below.

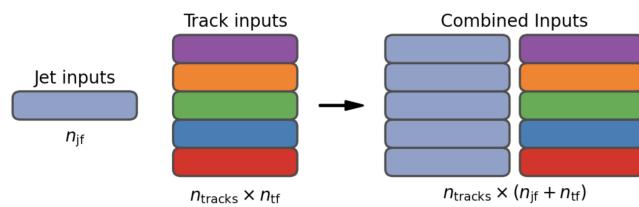


Figure 6.3: The inputs to GN1 are the two jet features ($n_{jf} = 2$), and an array of n_{tracks} , where each track is described by 21 track features ($n_{tf} = 21$) [2]. The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1.

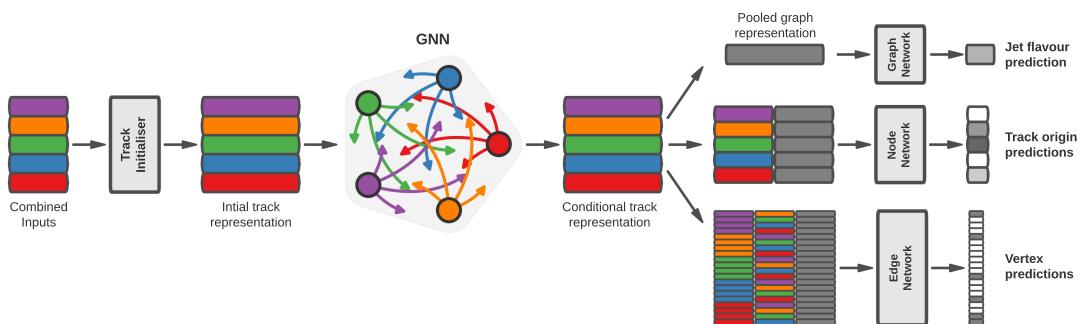


Figure 6.4: The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting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

1478 The outputs of the track initialisation network are used to populate the nodes of a
 1479 fully connected graph, such that each node in the graph neighbours every other node.
 1480 Each node h_i in the graph corresponds to a single track in the jet, and is characterised
 1481 by a feature vector, also called a representation. The per-track output representations
 1482 from the initialisation networks are used as the initial feature vectors of each node
 1483 in the graph. In each layer of the graph network, output node representations h'_i
 1484 are computed by aggregating the features of h_i and neighbouring nodes \mathcal{N}_i using
 1485 a multi-head attention mechanism ($n = 2$) as described in Ref. [119, 120]. First,
 1486 the feature vectors of receiver and sender nodes are fed into two fully connected
 1487 linear layers \mathbf{W}_r and \mathbf{W}_s , to produce an updated representation for each sender and
 1488 receiver node $\mathbf{W}_r h_i$ and $\mathbf{W}_s h_j$. These updated feature vectors are used to compute
 1489 edge scores $e(h_i, h_j)$ for each node pair, as in

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

1490 where, θ is a non-linear activation function, and \mathbf{a} is a learned vector. These edge
 1491 scores are then used to calculate attention weights a_{ij} for each pair of nodes using
 1492 the softmax function over the edge scores

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

1493 Finally, the updated representations for the receiver nodes h'_i are computed by taking
 1494 the weighted sum over each updated node representation $\mathbf{W}_r h_i$, with weights a_{ij}

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

1495 The set of operations described above constitute a single graph network layer. Three
 1496 such layers are stacked to construct the graph network, representing a balance
 1497 between achieving good performance in a reasonable time and avoiding overtraining

1498 due to inflation of the parameter count of the model. The final output from the graph
 1499 neural network is a set of per-node (i.e. per-track) feature vectors that are conditional
 1500 representations of each track given the other tracks in the jet. In order to perform
 1501 the jet flavour prediction, a flattened global representation of the jet is needed. To
 1502 produce this, the output track representations are combined using a weighted sum,
 1503 where the weights are learned during training and therefore act as a form of attention
 1504 over the different tracks. The flattened outputs from the sum are then fed into a
 1505 fully connected feedforward neural network with four layers and three outputs, one
 1506 for each jet flavour. Two other separate fully connected feedforward neural networks
 1507 are then also used to independently perform the auxiliary classification objectives of
 1508 GN1. Both of the auxiliary classification tasks also take in the global representation
 1509 of the jet as inputs. A summary of the different classification networks used for the
 1510 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. [2]. The hidden layers column contains a list specifying the number of neurons in each layer.

1511 The node classification network predicts the track truth origin as defined in Table 5.1.
 1512 This network takes as inputs the features from a single output node from the graph
 1513 network and the global representation of the jet. The node network has three hidden
 1514 layers containing 128, 64 and 32 neurons respectively, and an output size of seven,
 1515 corresponding to the seven different truth origins defined in Table 5.1.

1516 The edge classification network is used to predict whether the tracks in the track-
 1517 pair belong to a common vertex. This network takes as inputs the concatenated
 1518 representations from each pair of tracks and the global jet representation. Similar
 1519 to the node network, the edge network has three hidden layers containing 128,
 1520 64 and 32 neurons respectively, and a single output, which is used to perform
 1521 binary classification of the track-pair compatibility. The output predictions for the

1522 two auxiliary networks are used for the auxiliary training objectives discussed in
1523 Section 6.4.2.

1524 Finally, the graph classification network is used to predict the jet flavour. This
1525 network takes only the global jet representation as input. The graph classification
1526 network is comprised of four fully connected hidden layers with 128, 64, 32 and 16
1527 neurons respectively, and has three outputs corresponding to the b -, c - and light-jet
1528 classes.

1529 6.4.4 Training

1530 The full GN1 training procedure minimises the total loss function L_{total} , defined
1531 in Eq. (6.7). This loss is composed of three terms: L_{jet} , the categorical cross
1532 entropy loss over the different jet flavours; L_{vertex} , the binary track-pair compatibility
1533 cross entropy loss; and L_{track} , the categorical cross entropy loss for the track origin
1534 prediction. L_{vertex} is computed via a weighted average over all intra-jet track-pairs in
1535 the batch, and L_{track} is computed by a weighted average over all tracks in the batch,
1536 where the weights are described below.

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

1537 The different losses converge to different values during training, reflecting differences
1538 in the relative difficulty of the various training objectives. The values of L_{vertex} and
1539 L_{track} are weighted by $\alpha = 1.5$ and $\beta = 0.5$ respectively to ensure they converge to
1540 similar values, giving them an equal weighting towards L_{total} . The values of α and β
1541 are chosen to ensure that L_{jet} converges to a larger value than either L_{vertex} and L_{track} ,
1542 which reflects the primary importance of the jet classification objective. It was found
1543 that in practice the overall performance of the model was not sensitive to modest
1544 changes in the loss weights α and β . Pre-training using L_{total} (i.e. on all tasks) and
1545 fine tuning on only the jet classification task also did not improve performance versus
1546 the described standard setup, indicating that the auxiliary tasks are not in direct
1547 competition with the jet classification task. As there was a large variation in the
1548 relative abundance of tracks of the different origins, the contribution of each origin to

1549 L_{track} was weighted by the inverse of the frequency of their occurrence. In vertexing
1550 loss L_{vertex} , the class weight for track-pairs where both tracks are from either a b - or
1551 c -hadron was increased by a factor of two as compared with other track-pairs, to
1552 encourage the network to focus on correctly classifying heavy flavour vertices.

1553 GN1 can be trained with either the node or edge networks (and their corresponding
1554 auxiliary tasks), or both, removed, as discussed in Section 6.5.3. In such cases,
1555 the corresponding losses L_{vertex} and L_{track} are also removed from the calculation
1556 of the overall loss L_{total} . The performance of the resulting models provides a
1557 useful indication of the benefit of including the auxiliary tasks to the primary jet
1558 classification objective.

1559 GN1 was trained for 100 epochs on 4 NVIDIA V100 GPUs. Each epoch takes
1560 approximately 25 mins to complete over the training sample of 30 million jets
1561 described in Section 6.3.1. The Adam optimiser [121] with an initial learning rate of
1562 $1e-3$, and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically
1563 the validation loss, calculated on 500k jets, became stable after around 60 epochs.
1564 The epoch that minimized the validation loss was used for evaluation. GN1 has
1565 been integrated into the ATLAS software [53] using ONNX [122]. The test sample
1566 jet flavour predictions scores are computed using the ATLAS software stack as a
1567 verification of this process.

1568 6.5 Results

1569 The GN1 tagger is evaluated both as a b -tagging and c -tagging algorithm in Sec-
1570 tion 6.5.1 and Section 6.5.2 respectively. Evaluation is performed separately on
1571 both jets in the $t\bar{t}$ sample with $20 < p_{\text{T}} < 250 \text{ GeV}$ and jets in the Z' sample with
1572 $250 < p_{\text{T}} < 5000 \text{ GeV}$. The performance of the model is compared to the DL1r
1573 tagger [77, 111], which has been retrained on 75 million jets from the same samples
1574 as GN1. The input RNNIP tagger [79] to DL1r has not been retrained. As discussed,
1575 each tagger predicts the probability that a jet belongs to the b -, c - and light-classes.
1576 To use the model for b -tagging, these probabilities are combined into a single score
1577 D_b , which is defined as

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

1578 where f_c is a free parameter that determines the relative weight of p_c to p_l in the
 1579 score D_b , controlling the trade-off between c - and light-jet rejection performance.
 1580 The choice of f_c is arbitrary, and is optimised based upon the desired light- vs c -jet
 1581 rejection performance. This parameter is set to a value of $f_c = 0.018$ for the DL1r
 1582 model, obtained through an optimisation procedure described in Ref. [77]. Based on
 1583 a similar optimisation procedure, a value of $f_c = 0.05$ is used for the GN1 models.
 1584 A fixed-cut working point (WP) defines the corresponding selection applied to the
 1585 tagging discriminant D_b in order to achieve a given efficiency on the inclusive $t\bar{t}$
 1586 sample.

1587 A comparison of the b -tagging discriminant D_b between DL1r and GN1 is shown in
 1588 Fig. 6.5. The shapes of the D_b distributions are generally similar for b -, c - and light-
 1589 jets between both models, however, GN1 shifts the b -jet distribution to higher values
 1590 of D_b in the regions with the greatest discrimination. The GN1 c -jet distribution is
 1591 also shifted to lower values of D_b when compared with DL1r, enhancing the separation
 1592 and indicating that GN1 is improving c -jet rejection when compared with DL1r.

1593 6.5.1 b -tagging Performance

1594 The performance of b -tagging algorithms is quantified by their ability to reject c - and
 1595 light-jets for a given b -jet selection efficiency WP. In order to compare the b -tagging
 1596 performance of the different taggers for the b -jet tagging efficiencies in the range
 1597 typically used by analyses, the corresponding c - and light-jet rejection rates are
 1598 displayed in Figs. 6.6 and 6.7 for jets in the $t\bar{t}$ and Z' samples respectively. Four
 1599 standard WPs are defined with b -jet tagging efficiencies of 60%, 70%, 77% and 85%
 1600 respectively. These WPs are commonly used by physics analyses depending on their
 1601 specific signal and background requirements. The WPs are defined based on jets
 1602 in the $t\bar{t}$ sample only. Due to the much higher jet p_T range in the Z' sample, and
 1603 the increased difficulty in tagging jets at high- p_T (see Chapter 4), the b -jet tagging
 1604 efficiencies for jets in the Z' sample are lower than the corresponding WPs calculated
 1605 in the $t\bar{t}$ sample. For instance the WP cut value computed to provide a 70% b -jet

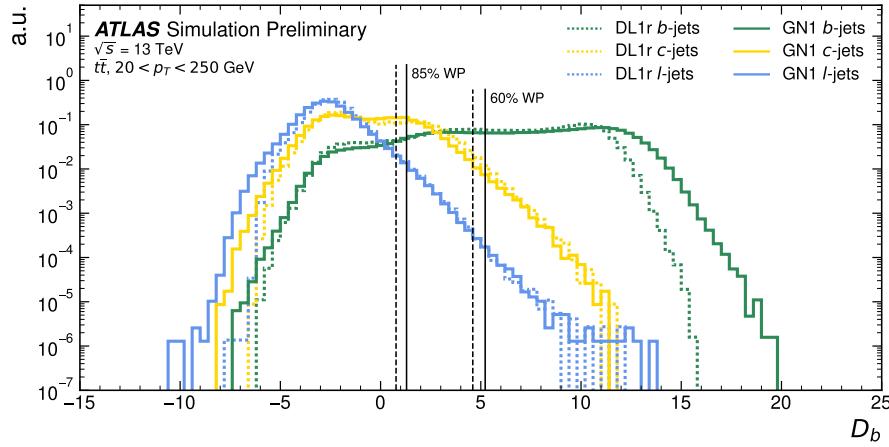


Figure 6.5: Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample [2]. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of $f_c = 0.018$ 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.

1606 tagging efficiency on the $t\bar{t}$ sample results in a b -jet tagging efficiency of just $\sim 30\%$
 1607 on the Z' sample. In order to account for this, the range of b -jet tagging efficiencies
 1608 displayed for plots showing the performance for jets in the Z' sample (for example
 1609 Fig. 6.7) is chosen to span the lower efficiencies achieved in the Z' sample at high- p_T .

1610 For jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV, GN1 demonstrates considerably
 1611 better c - and light-jet rejection when compared with DL1r across the full range of
 1612 b -jet tagging efficiencies studied. The relative improvement is strongly dependent
 1613 on the b -jet tagging efficiency under study. The largest improvements are found at
 1614 lower b -jet tagging efficiencies. At a b -jet tagging efficiency of 70%, the c -jet rejection
 1615 improves by a factor of ~ 2.1 while the light-jet rejection improves by a factor of ~ 1.8
 1616 with respect to DL1r. For high- p_T jets in the Z' sample with $250 < p_T < 5000$ GeV,
 1617 GN1 also brings a significant performance improvement with respect to DL1r across
 1618 the range of b -jet tagging efficiencies studied. Again, the largest relative improvement
 1619 in performance comes at the lower b -jet tagging efficiencies. At a b -jet efficiency of
 1620 30%, GN1 improves the c -jet rejection with respect to DL1r by a factor of ~ 2.8 and
 1621 the light-jet rejection by a factor of ~ 6 . The performance comparison at lower b -jet
 1622 tagging efficiencies is made more difficult due to the increased statistical uncertainties
 1623 which result from the high rejection of background. Given that GN1 exploits the

1624 low-level detector information in a more complete and sophisticated way than DL1r,
1625 further studies are needed to confirm that the performance gain observed in these
1626 simulated samples is also observed in experimental data.

1627 The GN1Lep variant of GN1 demonstrates further improved performance with respect
1628 to the baseline model. This demonstrates the additional jet flavour discrimination
1629 power provided by the leptonID track input. For jets in the $t\bar{t}$ sample, the relative c -
1630 jet rejection improvement with respect to GN1 at the 70% b -jet WP is approximately
1631 30%. The improvement in light-jet rejection also increases by 40% at the same WP.
1632 For jets in the Z' sample, the relative c -jet rejection (light-jet rejection) performance
1633 with respect to GN1 improves by approximately 10% (25%) at a b -jet tagging
1634 efficiency of 30%.

1635 In general, the performance of all the taggers is strongly dependent on the jet p_T .
1636 This is due to the increased multiplicity and collimation of tracks, and the displaced
1637 decays that result from within the heavy flavour jets (see Chapter 4). Together,
1638 they contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a
1639 general degradation in quality of tracks inside the core of a jet, which in turn reduces
1640 the jet tagging performance.

1641 In order to study how the tagging performance changes as a function of the jet p_T ,
1642 the b -jet tagging efficiency as a function of p_T for a fixed light-jet rejection of 100 in
1643 each bin is shown in Fig. 6.8. For jets in the $t\bar{t}$ sample, at a fixed light-jet rejection
1644 of 100, GN1 improves the b -jet tagging efficiency by approximately 4% across all the
1645 jet p_T bins. Meanwhile, GN1Lep again demonstrates improved performance with
1646 respect to GN1, in particular at lower p_T . The relative increase in the b -jet tagging
1647 efficiency increases from 4% to 8% with respect to DL1r. For jets in the Z' sample,
1648 GN1 again outperforms DL1r across the entire jet p_T range studied. The largest
1649 relative improvement in performance is found at the highest transverse momenta
1650 of jet $p_T > 2 \text{ TeV}$, and corresponds to an approximate factor of 2 improvement in
1651 efficiency with respect to DL1r.

1652 The performance of the model was also evaluated as a function of the average
1653 number of pileup interactions in the event. No significant dependence of the tagging
1654 performance was observed.

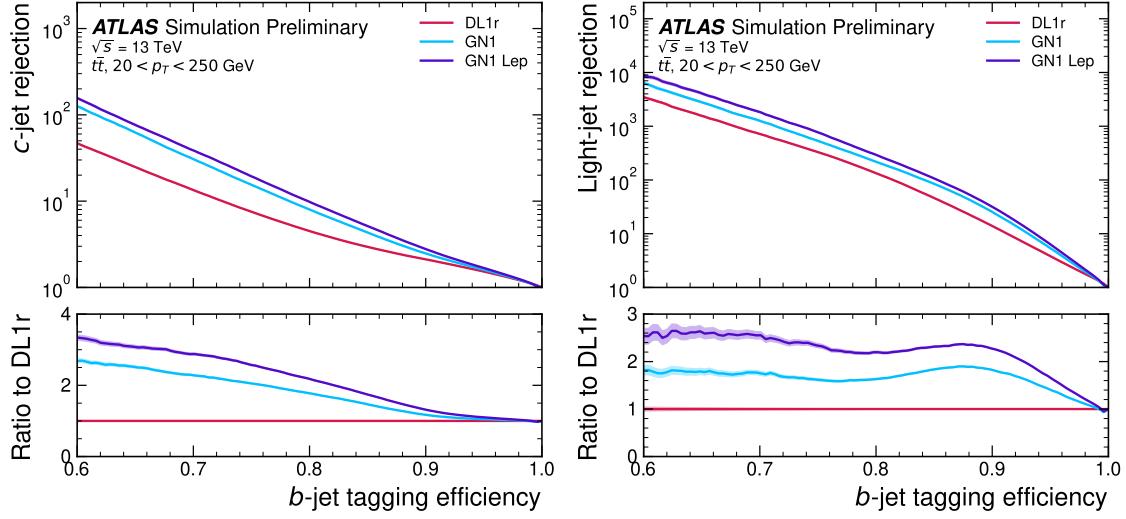


Figure 6.6: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$ [2]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ 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.

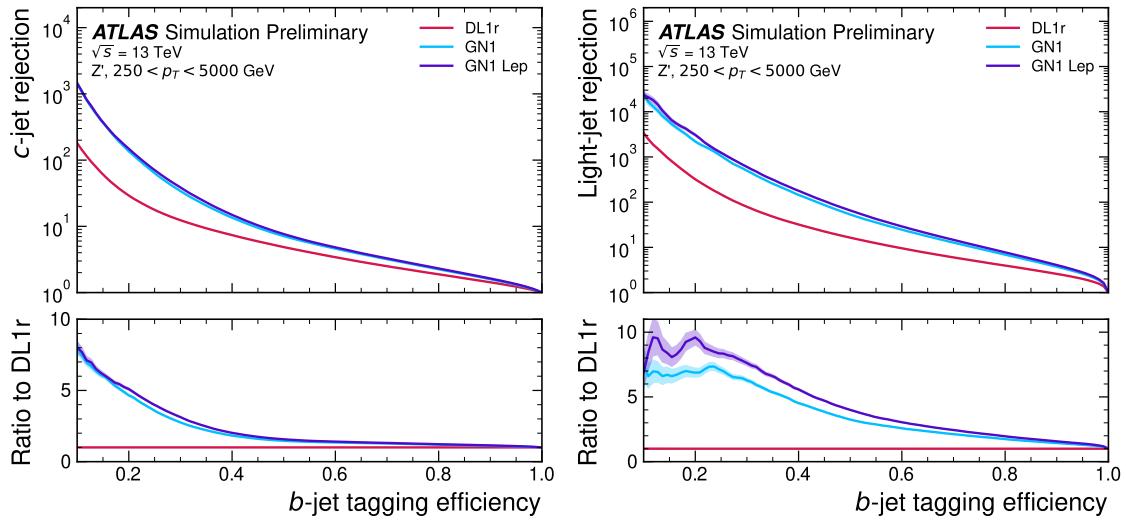


Figure 6.7: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$ [2]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used 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.

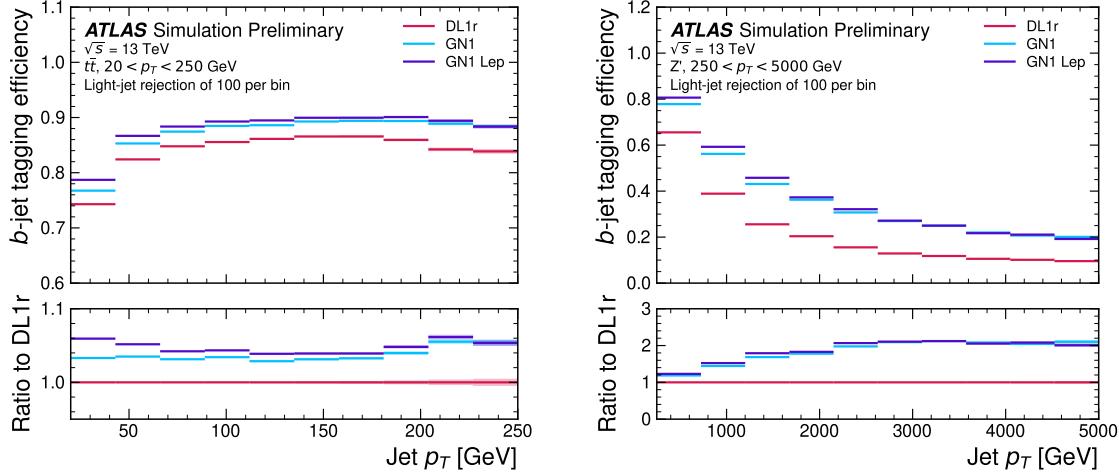


Figure 6.8: The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin [2]. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. GN1 demonstrates improved performance with respect to DL1r across the p_T range shown. Binomial error bands are denoted by the shaded regions.

6.5.2 c -tagging Performance

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

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

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. As in Section 6.5.1,

1668 the WPs are defined using jets in the $t\bar{t}$ sample. Standard c -jet tagging efficiency
 1669 WPs used by physics analyses are significantly lower than the b -tagging WPs in order
 1670 to maintain reasonable b - and light-jet rejection rates. This is reflected in the range
 1671 of c -jet tagging efficiencies used in c -tagging plots such as Figs. 6.9 and 6.10. Fig. 6.9
 1672 displays the c -tagging performance of the models on the jets in the $t\bar{t}$ sample. GN1
 1673 is shown to perform significantly better than DL1r. Similar to the b -tagging case,
 1674 the b - and light-jet rejection improve most at lower c -jet tagging efficiencies, with
 1675 the c -jet rejection (light-jet rejection) improving by a factor 2 (1.6) with respect to
 1676 DL1r at a c -jet tagging efficiency of 25%. GN1Lep again outperforms GN1, though
 1677 the improvements are more modest than observed for the b -tagging use case, with
 1678 the both the b -jet rejection and light-jet rejection improving with respect to GN1 by
 1679 approximately 10% at the 25% c -jet WP. Fig. 6.10 shows the c -tagging performance
 1680 on the jets in the Z' sample with $250 < p_T < 5000$ GeV. Both GN1 and GN1Lep
 1681 perform similarly, improving the b -jet rejection by 60% and the light-jet rejection by
 1682 a factor of 2 at the 25% c -jet WP.

1683 6.5.3 Ablations

1684 Ablation studies (the removal of certain components of a given model in order to
 1685 study the impact of that component) are carried out to determine the
 1686 importance of the auxiliary training objectives of GN1 to the overall performance.
 1687 The “GN1 No Aux” variant retains the primary jet classification objective, but
 1688 removes both track classification and vertexing auxiliary objectives (see Section 6.4.2)
 1689 and correspondingly only minimises the jet classification loss. The “GN1 TC” variant
 1690 includes track classification objective but not the vertexing objective. Finally, the
 1691 “GN1 Vert” includes the vertexing objective, but not the track classification objective.

1692 For jets in both the $t\bar{t}$ and Z' samples, a general trend is observed that the models
 1693 trained without one or both of the auxiliary objectives results in significantly reduced
 1694 c - and light-jet rejection when compared with the baseline GN1 model. This result
 1695 is shown clearly in Figs. 6.11 and 6.12. For jets in the $t\bar{t}$ sample, the performance of
 1696 GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert perform similarly to
 1697 each other. For jets in the Z' sample meanwhile, the GN1 No Aux model already
 1698 shows a clear improvement in c - and light-jet rejection when compared with DL1r at
 1699 lower b -jet tagging efficiencies. Similar to jets in the $t\bar{t}$ sample, GN1 TC and GN1

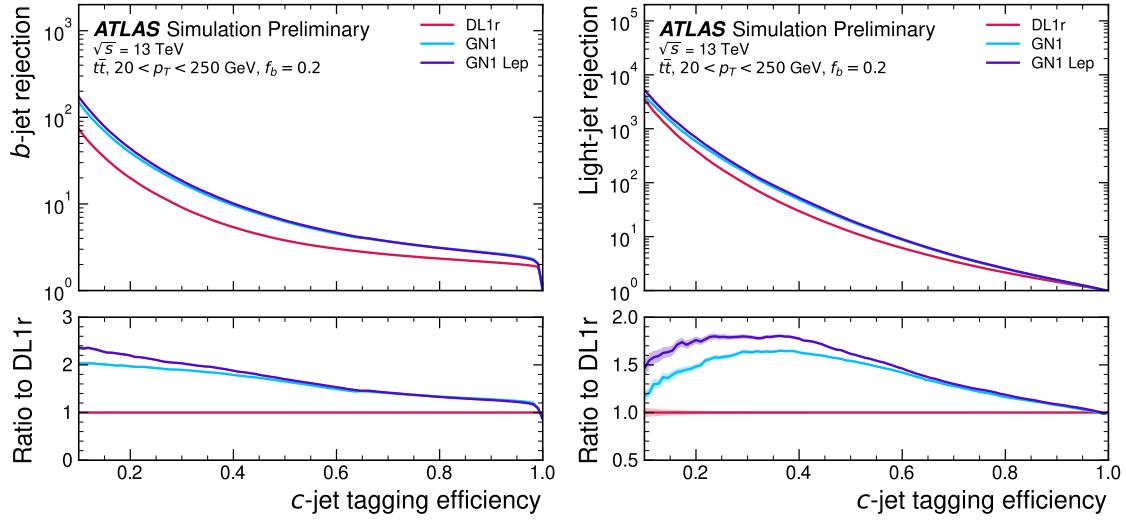


Figure 6.9: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$ [2]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.

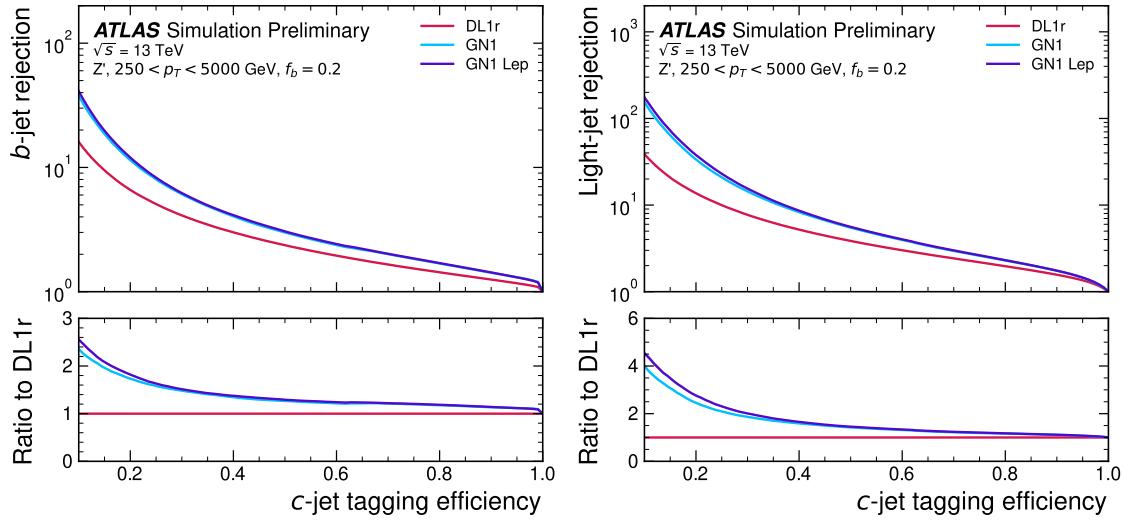


Figure 6.10: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000 \text{ GeV}$ [2]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.

1700 Vert perform similarly, and bring large gains in background rejection when compared
1701 with GN1 No Aux, but the combination of both auxiliary objectives yields the best
1702 performance.

1703 It is notable that the GN1 No Aux model matches or exceeds the performance of
1704 DL1r without the need for inputs from the low-level algorithms. This indicates that
1705 the performance improvements enabled by the improved neural network architecture
1706 used in GN1 appear to be able to compensate for the removal of the low-level
1707 algorithm inputs. The GN1 TC and GN1 Vert variants each similarly outperform
1708 DL1r, demonstrating that both contribute to the overall high performance of the
1709 baseline model. The overall best performing model is the full version of GN1 trained
1710 with both auxiliary objective, demonstrating that the two auxiliary objectives are
1711 complementary.

1712 6.5.4 Inclusion of Low-Level Vertexing Algorithms

1713 As already mentioned, GN1 does not include any inputs from the low-level tagging
1714 algorithms, including the vertexing algorithms SV1 and JetFitter [61]. Since these
1715 algorithms are known to play a key role in contributing to the performance of DL1r, it
1716 was necessary to study whether their inclusion in GN1 resulted in further performance
1717 improvements. In a dedicated training of GN1 the SV1 and JetFitter tagger outputs
1718 were added to the GN1 jet classification network as an input, similar to how they
1719 are used in DL1r. These outputs include information on the reconstructed vertices,
1720 including the number of vertices, and the properties of the highest ranking recon-
1721 structed vertex (in the case of JetFitter). In addition, the index of the reconstructed
1722 SV1 or JetFitter vertices were included as two track-level inputs to GN1. These
1723 indices were also used to construct an an input feature for the edge classification
1724 network used to identify vertices, which was given a value of one if the track-pair
1725 were from a common reconstructed SV1 or JetFitter vertex, and zero otherwise. The
1726 jet classification performance of this GN1 model was not significantly different to the
1727 baseline model, and in some cases the performance was slightly reduced. GN1 does
1728 not benefit from the inclusion of information from SV1 and JetFitter, indicating
1729 that the model is able to reconstruct the relevant information provided by these
1730 low-level algorithms. The study also demonstrates that the model can function as a
1731 highly performant standalone tagger that does not require (beyond retraining) any

determine
ranking

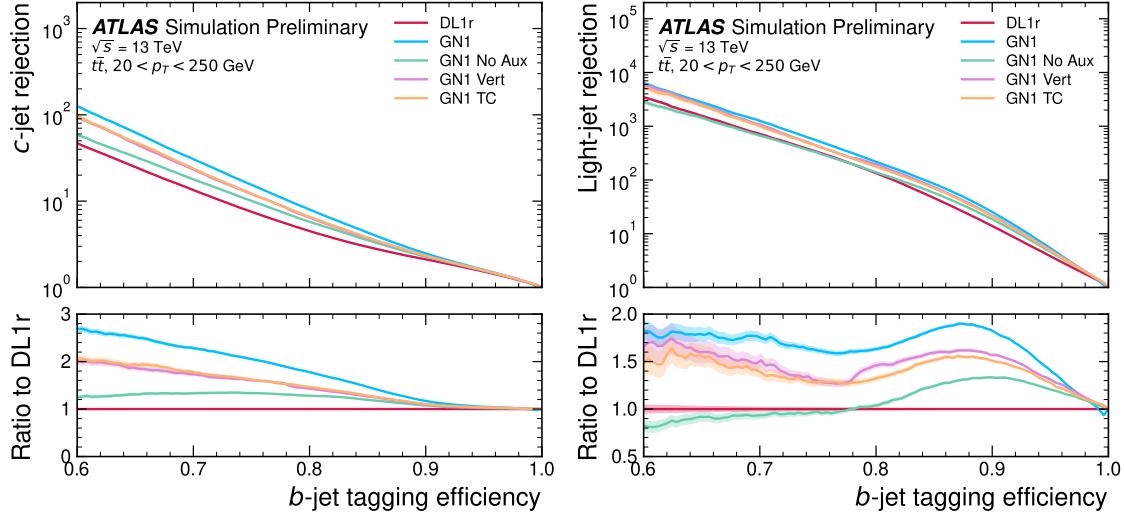


Figure 6.11: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed [2]. 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.

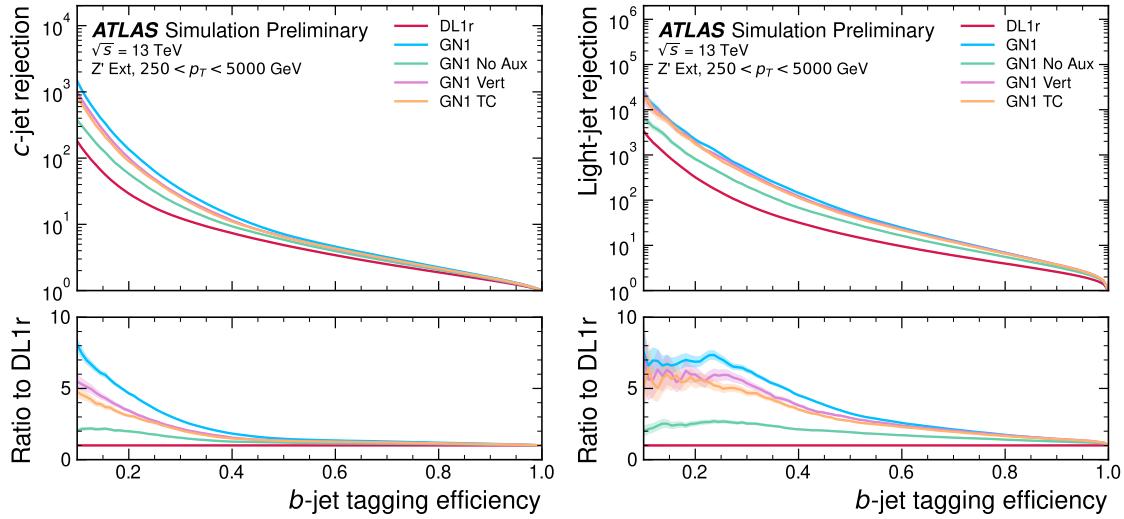


Figure 6.12: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed [2]. 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.

1732 manual optimisation to achieve good performance in a wide range of phase spaces.
1733 A dedicated look at the vertexing performance of GN1 with some comparisons to
1734 SV1 and JetFitter is found in Section 6.5.6

1735 6.5.5 Jet Display Diagrams

1736 The auxiliary training objectives of GN1 allow for improved model interpretability,
1737 which is especially important for a monolithic approach as the low level taggers,
1738 which provide useful physical insight, are no longer present. Figs. 6.13 and 6.14
1739 provide example comparisons of the true origin and vertexing information compared
1740 with the predicted values from GN1, SV1 and JetFitter. Such comparisons can be
1741 used to provide an indication that GN1 reconstructs the correct representation of the
1742 jet structure, and may also help to identify limitations of the model. In the figures,
1743 the tracks in the jet are indexed twice on each of the x - and y -axes, and tracks are
1744 grouped into vertices along with other tracks as indicated by common markings in
1745 the relevant rows and columns.

1746 In Fig. 6.13, GN1 correctly groups the three primary tracks as having come from the
1747 primary vertex. The b -hadron and $b \rightarrow c$ -hadron decay vertices are also correctly
1748 predicted, and the origin of the tracks in each is correct. There is a single OtherSec-
1749 ondary track which GN1 incorrectly predicts as having come pileup. Meanwhile SV1
1750 (by design) merges the two heavy flavour decay vertices, but incorrectly includes a
1751 track from the primary vertex. JetFitter reconstructs two vertices, one which is a
1752 combination of two tracks from different truth vertices and two other single track
1753 vertices in each of the heavy flavour vertices. GN1 also predicts the flavour of the jet
1754 with a high degree of certainty.

1755 Similarly, Fig. 6.13 shows that GN1 is able to relatively accurately predict the origin
1756 and vertex information of tracks inside a jet. The pileup tracks and primary vertex
1757 tracks are correctly identified, and the heavy flavour decay tracks are also correctly
1758 identified with the exception of one of the b -hadron decay tracks. Again, SV1 merges
1759 the two heavy flavour decay vertices along with a track from pileup, while JetFitter
1760 shows signs of being underconstrained by reconstructing two single track vertices,
1761 one with a pileup track and one with a track from a $b \rightarrow c$ -hadron decay.

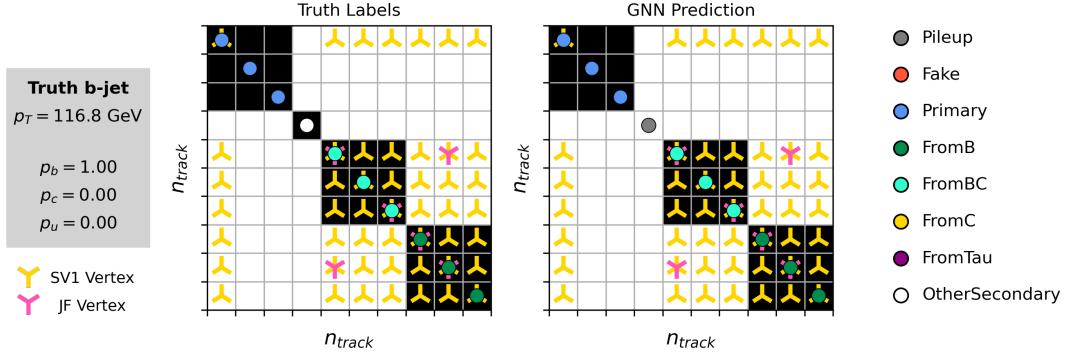


Figure 6.13: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed 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 pileup.

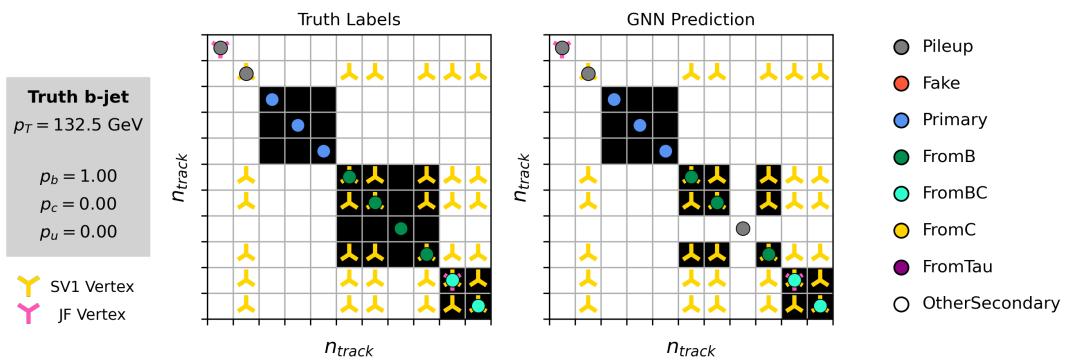


Figure 6.14: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a b -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed 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.

6.5.6 Vertexing Performance

From the track-pair vertex prediction described in Section 6.4.2, tracks can be partitioned into compatible groups representing vertices (see [113]). As such, GN1 can perform vertex “finding”, but not vertex “fitting”, i.e. the reconstruction of a vertex’s properties, which currently still requires the use of a dedicated vertex fitter. In order to study the performance of the different vertexing tools, the truth vertex label of the tracks, discussed in Section 6.4.2, are used. To estimate the efficiency with which GN1 manages to find vertices inclusively, vertices containing tracks identified as coming from a b -hadron are merged together and compared to the inclusive truth decay vertices that result from a b -hadron decay (where if there are multiple distinct truth vertices from a b -hadron decay they are also merged together). Vertices are compared with the target truth vertex and the number of correctly and incorrectly assigned tracks is computed. Since secondary vertex information is only recovered for reconstructed tracks, a vertex finding efficiency of 100% denotes that all possible secondary vertices are found given the limits set by the track reconstruction efficiency. A vertex is considered matched if it contains at least 65% of the tracks in the corresponding truth vertex, and has a purity of at least 50%. GN1 manages to achieve an inclusive reconstruction efficiency in b -jets of $\sim 80\%$, demonstrating that it effectively manages to identify the displaced vertices from b -hadron decays.

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

1795 design constraints. The inclusive vertex finding performance of all three tools can be
1796 compared using the procedure outlined below.

1797 The starting point for the secondary vertex finding efficiency in both the exclusive and
1798 inclusive cases is to select truth secondary vertices, defined as those containing only
1799 inclusive b -hadron decays. For exclusive vertex finding, these truth secondary vertices
1800 can be used directly as the denominator for the efficiency calculation. Meanwhile
1801 for the inclusive efficiency all such truth secondary vertices in the jet are merged
1802 into a single inclusive target vertex. Correspondingly, for the inclusive vertex finding
1803 case, the vertices found by JetFitter are merged into a single vertex, and the vertices
1804 found by GN1, which contain at least one predicted b -hadron decay track, are also
1805 merged similarly. SV1 does not require any vertex merging. Only jets containing a
1806 single b -hadron at truth level are considered.

1807 Next, vertices in the jet found by the different vertexing tools are compared with
1808 the target truth vertices. The number of correctly and incorrectly assigned tracks is
1809 computed. In order to call a vertex efficient, it is required to contain at least 65% of
1810 the tracks in the corresponding truth vertex, and to have a purity of at least 50%.
1811 Single track vertices are required to have a purity of 100%. Additionally, for GN1
1812 only, at least one track in the vertex is required to have a predicted heavy flavour
1813 origin.

1814 Vertex finding efficiencies for b -jets in the $t\bar{t}$ sample are displayed as a function of p_T
1815 separately for the inclusive and exclusive approaches in Fig. 6.15. For b -jets in the $t\bar{t}$
1816 sample with $20 < p_T < 250$ GeV, the exclusive vertex finding efficiency of JetFitter
1817 and GN1 is relatively flat as a function of p_T . For the truth secondary vertices in this
1818 p_T region, JetFitter efficiently finds approximately 40% and GN1 finds approximately
1819 55%. When finding vertices inclusively the vertex finding efficiency is generally higher.
1820 An increased dependence on p_T is also visible for JetFitter and SV1. As the jet p_T
1821 increases from 20 GeV to 100 GeV, the efficiency of JetFitter increases from 60% to
1822 65%. In the same range, the efficiency of SV1 increases from 60% to 75%. GN1
1823 displays less dependence on p_T than JetFitter and SV1, efficiently finding upwards
1824 of 80% of vertices in b -jets in this p_T region. For b -jets with $p_T > 100$ GeV, JetFitter
1825 finds approximately 65% of vertices, SV1 finds approximately 75% of vertices, and
1826 GN1 finds approximately 80% of vertices.

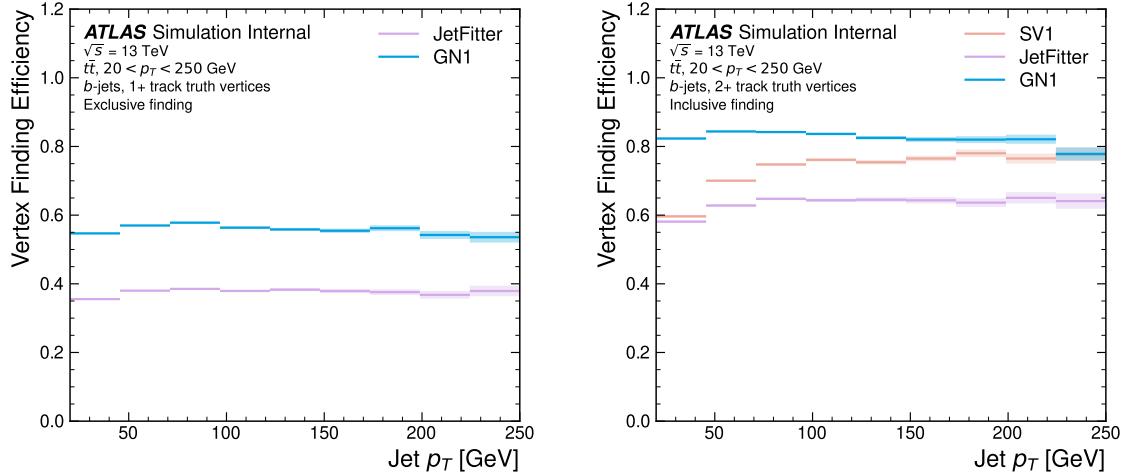


Figure 6.15: Vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at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

Fig. 6.16 compares the exclusive vertex finding efficiencies of JetFitter and GN1 for multitrack vertices. For b -jets in the Z' sample, the 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 75%, while the efficiencies of SV1 and JetFitter are around 60%. The efficiency of SV1 drops rapidly to almost zero above 3 TeV, while JetFitter and GN1 retain approximately 25% and 30% efficiency respectively. JetFitter finds 45-50% of vertices in b -jets in the $t\bar{t}$ sample, while GN1 finds 60-65%. For b -jets in the Z' sample, 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.

6.5.7 Track Classification Performance

One of the two auxiliary training objectives used by GN1 is to predict the truth origin of each track associated to the jet, as discussed in Section 6.4.2. Since the equivalent information is not provided by any of the existing flavour tagging tools, a benchmark model used to predict the truth origin of each track is trained based on a standard

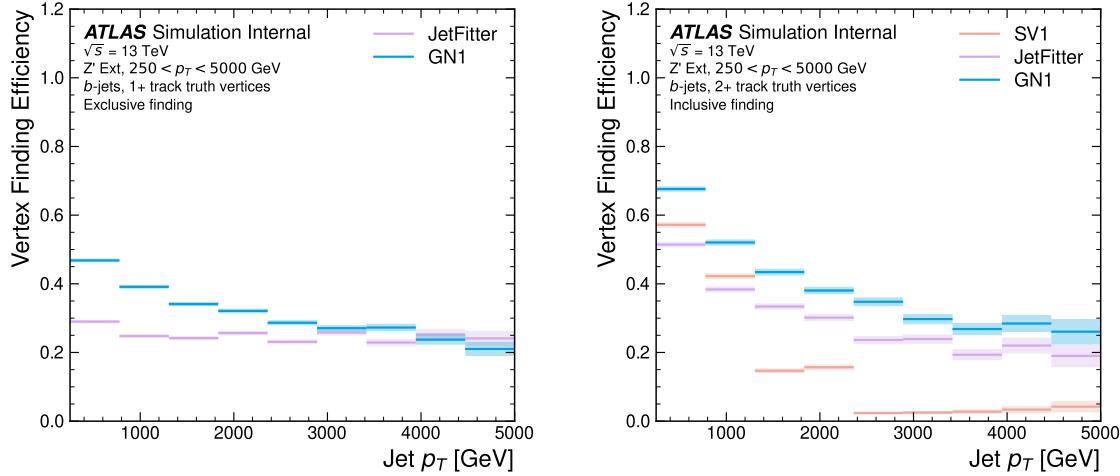


Figure 6.16: Inclusive vertex finding efficiency for multitrack truth vertices in b -jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T . The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain 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.

1842 multi-class feed-forward classification network. The benchmark model is trained
 1843 on the same tracks used for the baseline GN1 training. The model uses precisely
 1844 the same concatenated track-and-jet inputs as used by GN1 (see Section 6.4.1),
 1845 but processes only a single track at a time, meaning it cannot take into account
 1846 the correlations between tracks when determining the track origin. The model is
 1847 made up of five densely connected linear layers with 200 neurons in each layer. The
 1848 performance of the model was found to be unsensitive to changes in the network
 1849 structure.

1850 To measure the track classification performance, the area under the curve (AUC)
 1851 of the receiver operating characteristic (ROC) curve is computed for each origin
 1852 class, using a one-versus-all classification approach. The AUCs for the different
 1853 truth origins are averaged using both an unweighted and a weighted mean. The
 1854 unweighted mean treats the performance of each class equally, while the weighted
 1855 mean uses as a weight the relative abundance of tracks of each class. Table 6.3
 1856 demonstrates clearly that GN1 outperforms the MLP both at $20 < p_T < 250 \text{ GeV}$
 1857 for jets in the $t\bar{t}$ sample and at $250 < p_T < 5000 \text{ GeV}$ for jets in the Z' sample. For
 1858 example, GN1 can reject 65% of fake tracks in jets in the $t\bar{t}$ sample, while retaining

more than 99% of good tracks (i.e. those tracks which are not fake). The GN1 model has two advantages over the MLP which can explain the performance improvement. Firstly, the graph neural network architecture enables the sharing of information between tracks as discussed in Section 6.4.3. This is likely to be beneficial since the origins of different tracks within a jet are correlated. Secondly, the jet classification and vertexing objectives may be complementary to the track classification objective, and so the track classification performance is improved by the combined training of complementary objectives.

		AUC		Precision		Recall		F1	
		Mean	Weighted	Mean	Weighted	Mean	Weighted	Mean	Weighted
$t\bar{t}$	MLP	0.87	0.89	0.39	0.71	0.51	0.56	0.36	0.62
	GN1	0.92	0.95	0.51	0.82	0.64	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.

Fig. 6.17 shows the track origin classification ROC curves for the different track origins for jets in both the $t\bar{t}$ and Z' samples. In order to improve visual readability of the plot, the curves for the heavy flavour truth origins (FromB, FromBC and FromC) have been combined (weighted by their relative abundance), as have the Primary and OtherSecondary origins. In jets in both the $t\bar{t}$ and Z' samples, the AUC of all the different origin groups exceeds 0.9, representing strong overall classification performance. In both samples fake tracks are the easiest to classify, followed by pileup tracks. The FromC tracks which are c -hadron decay products, are the hardest to classify, possibly due to their similarity to both fragmentation tracks and b -hadron decay tracks, depending on the c -hadron species in question.

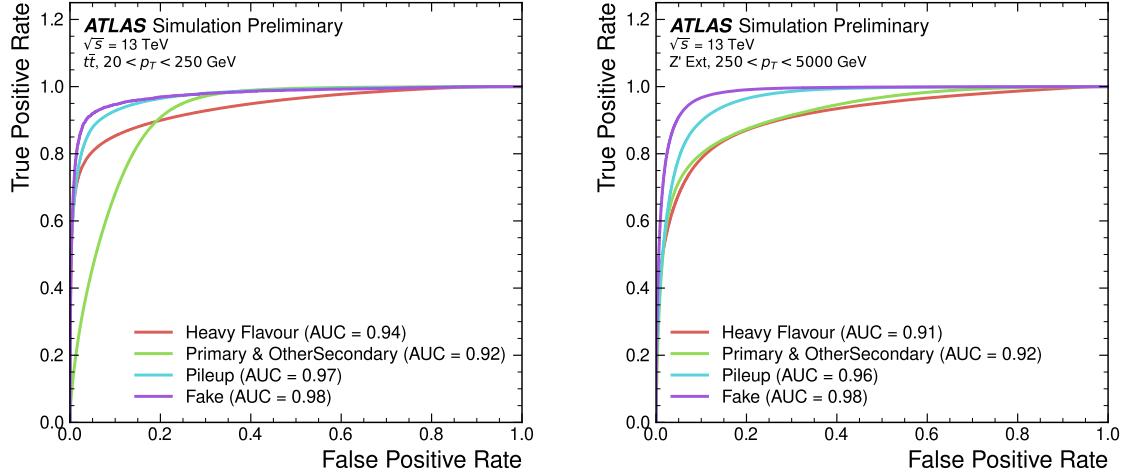


Figure 6.17: ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right). The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the 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.

1877 6.5.8 Looser Track Selection

1878 The track selections used to produce the main results are listed in Table 5.3. This
 1879 selection includes a cut on the number of shared silicon modules used to reconstruct
 1880 the track $N_{\text{shared}}^{\text{Si}}$. This value is calculated as

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

1881 where $N_{\text{shared}}^{\text{Pix}}$ is the number of shared pixel hits and $N_{\text{shared}}^{\text{SCT}}$ is the number of shared
 1882 SCT modules on a track. The nominal cut used elsewhere in this thesis is $N_{\text{shared}}^{\text{Si}} < 2$.
 1883 As the rate of shared hits is significantly higher for b -hadron decay tracks than for
 1884 other tracks, especially at high- p_T , this cut rejects a significant proportion of these
 1885 tracks. Figs. 6.18 and 6.19 show the result of training the GN1 tagger with the full
 1886 relaxation of this cut, i.e. allowing tracks with any number of shared hits. The shared
 1887 hit requirements applied by the ambiguity solver as part of track reconstruction (see
 1888 Section 3.4.1) are still applied. In addition, the maximum allowed value of d_0 is
 1889 increased from 3.5 mm to 5.0 mm. The results show that optimisation of the input
 1890 track selection can lead to significant improvements in performance over the default

selection. It is also possible that additional studies on further loosening the selection could yield further improved results. The maximum number of tracks provided as input to the model could also be increased from the default value of 40. In order to change the default tracking setup, studies investigating the modelling uncertainties of the additional tracks need to be carried out.

6.6 Implementations of GN1

- trigger plot
- upgrade plot

6.7 Conclusion

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

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

GN1 demonstrates improved track classification performance when compared with a simple per-track MLP and an efficiency of $\sim 80\%$ for inclusive vertex finding in

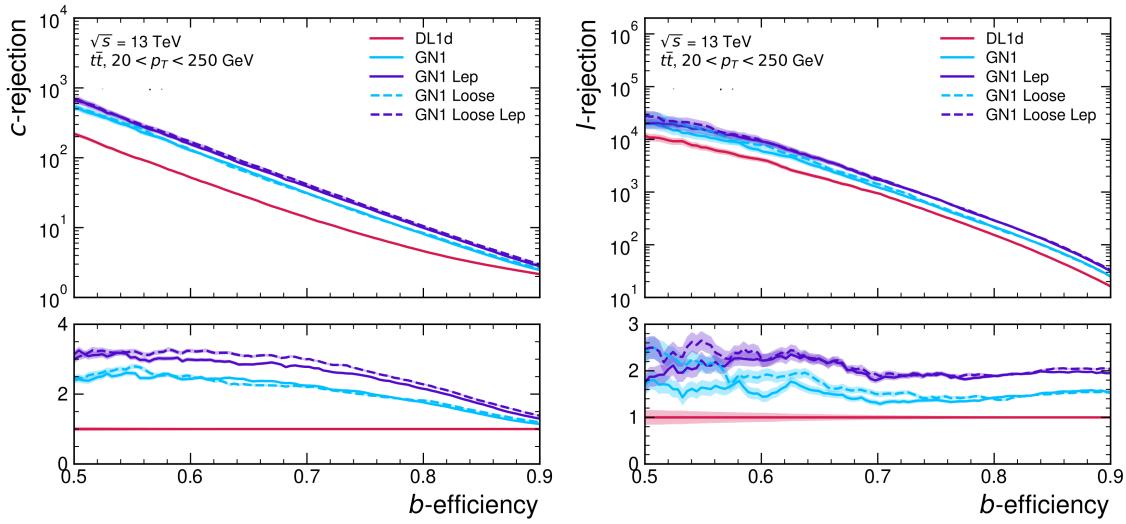


Figure 6.18: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions.

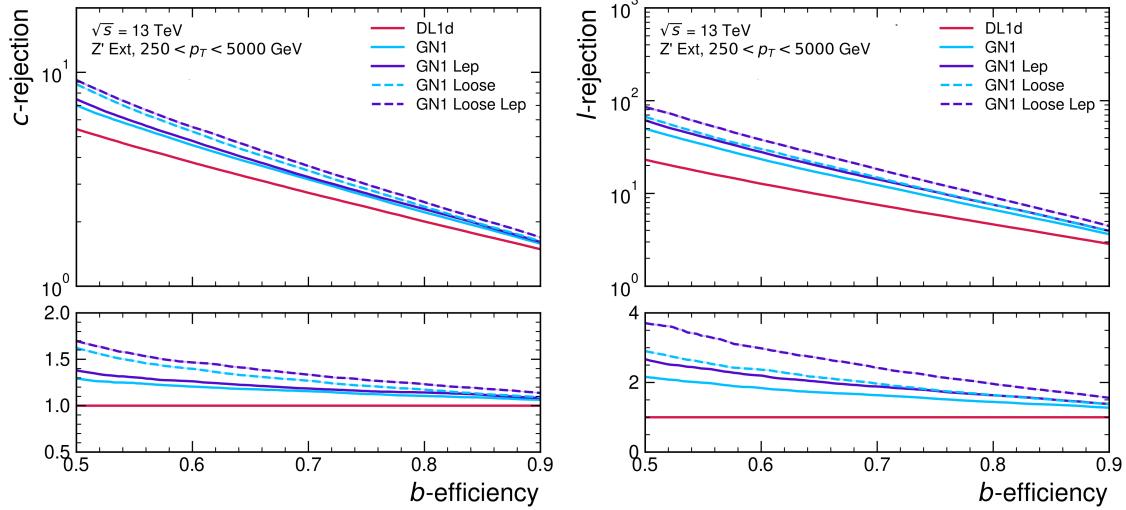


Figure 6.19: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions.

1919 b -jets. The model is also able to perform vertex finding, and preliminary studies
1920 suggest it outperforms previous manually optimised approaches. The auxiliary track
1921 classification and vertex finding objectives are shown to significantly contribute to
1922 the performance in the jet classification objective, and, along with the more advanced
1923 graph neural network architecture, are directly responsible for the improvement over
1924 DL1r.

1925 Further improvements in the b - and c -tagging performance are likely possible with
1926 a more thorough optimisation of the model architecture, and the integration of
1927 additional information from other parts of the ATLAS detector. The addition of
1928 other auxiliary training objectives, such as the truth b -hadron decay radius and
1929 transverse momentum, may also yield additional performance gains on top of the
1930 gains achieved by loosening the input track selection (demonstrated in Section 6.5.8).

1931 Additional future work includes the verification of the performance of GN1 on collision
1932 data, and the full calibration of the model so it can be used by analyses. The flexible
1933 nature of the model means it can also be readily applied to other related problems
1934 outside of standard b - and c -tagging applications, for example $X \rightarrow bb$ and $X \rightarrow cc$
1935 tagging. The model could also be repurposed as a pileup jet tagger, or general
1936 primary and secondary vertexing tool.

1937 The model also demonstrates strong performance for the HLLHC ??.

₁₉₃₈ **Chapter 7**

₁₉₃₉ **Boosted VHbb Analysis**

₁₉₄₀ The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [13, 14],
₁₉₄₁ is predicted by the standard model to decay primarily to a pair of b -quarks, with a
₁₉₄₂ branching factor of 0.582 ± 0.007 for $m_H = 125$ GeV [25]. Observation of this decay
₁₉₄₃ mode was reported by ATLAS [108] and CMS [27] in 2018, establishing the first
₁₉₄₄ direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks
₁₉₄₅ (see Section 2.2.2). The $H \rightarrow b\bar{b}$ process is also important for constraining the total
₁₉₄₆ decay width of the Higgs [123].

₁₉₄₇ Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion
₁₉₄₈ as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet
₁₉₄₉ background and so overall sensitivity to the Higgs is low. The QCD multijet
₁₉₅₀ background refers to events containing one or more strongly produced jets which are
₁₉₅₁ not the decay product of heavy resonances, for example $g \rightarrow q\bar{q}$. The gluon-gluon
₁₉₅₂ fusion channel contains to leading order only jets in the final state, and therefore
₁₉₅₃ it is extremely difficult to distinguish signal events from the overwhelming multijet
₁₉₅₄ background. The $H \rightarrow b\bar{b}$ observation therefore searched for Higgs bosons produced
₁₉₅₅ in association with a vector boson V (where V can be a W or Z boson) which
₁₉₅₆ subsequently decays leptonically. The leptonic final states from the decay of the
₁₉₅₇ vector boson allow for leptonic triggering whilst at the same time significantly
₁₉₅₈ reducing the multijet background.

₁₉₅₉ A closely related analysis [124] has more recently measured the associated production
₁₉₆₀ of a Higgs boson decaying to b -quarks in events where the vector and Higgs bosons are
₁₉₆₁ highly boosted. The analysis is outlined in Section 7.1. Modelling studies performed
₁₉₆₂ by the author are detailed in Section 7.2, and the results of the analysis are presented

1963 in Section 7.4. The author contributed to various signal and background modelling
1964 studies, fit studies, and to the diboson unblinding effort. This analysis has been
1965 published in Ref. [124]. Figures and tables from Ref. [124] are reproduced here.

1966 **7.1 Analysis Overview**

1967 The boosted VH , $H \rightarrow b\bar{b}$ analysis is focused on the high transverse momentum
1968 regime, which has the benefit of being more sensitive to physics beyond the Standard
1969 Model [125], but the disadvantage of being more challenging due to the increased dif-
1970 ficulty in the accurate reconstruction of highly energy events (discussed in Chapter 4).
1971 In order to focus on the high- p_T regime, the reconstructed vector boson is required to
1972 have $p_T^V > 250$ GeV (see Section 7.1.2). Events are also split into two p_T^V bins with the
1973 first bin covering $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and the second covering $p_T^V > 400 \text{ GeV}$,
1974 which allows the analysis to account for the improved signal-to-background in the
1975 high- p_T regime.

1976 The previous ATLAS analysis in Ref. [108] was primarily sensitive to vector bosons
1977 with a more modest p_T^V boost in the region of 100–300 GeV. In this regime, the Higgs
1978 candidate was reconstructed using a pair of jets with radius parameter of $R = 0.4$,
1979 called small- R jets. However in the high- p_T regime, the decay products of the Higgs
1980 boson become increasingly collimated and the small- R jets may overlap. In order to
1981 avoid the associated problems and to aid in the reconstruction of the Higgs boson
1982 candidate, the present analysis uses instead a large- R jet with radius parameter
1983 $R = 1.0$ to reconstruct the Higgs boson candidate in all channels (see Section 3.4.3).
1984 The Higgs candidate is required to have exactly two ghost-associated and b -tagged
1985 variable-radius track-jets. The candidate large- R jet is reconstructed using jet
1986 substructure techniques, for example it is trimmed by removing soft and wide-angle
1987 components, which helps to remove particles from the underlying event and pileup
1988 collisions [126]. Refer to Section 3.4.3 for more details on jet reconstruction.

1989 On top of the binning in p_T^V , selected events are further categorised into 0-, 1- and
1990 2-lepton channels depending on the number of selected charged leptons (electrons
1991 and muons) are present in the reconstructed final state (also referred to as 0L, 1L,
1992 and 2L respectively). The 0-lepton channel targets the $ZH \rightarrow \nu\nu b\bar{b}$ process, the
1993 1-lepton channel targets $WH \rightarrow \ell\nu b\bar{b}$, and the 2-lepton channel targets $ZH \rightarrow \ell\ell b\bar{b}$,

where ℓ is an electron or muon and ν is a neutrino. Each channel has a dedicated set of selections which are listed in more detail in Section 7.1.3. Events in the 0- and 1-lepton channels are further split depending on the number of additional small- R jets not matched to the Higgs-jet candidate. The high-purity signal region (HP SR) has zero such jets, while the low-purity signal region (LP SR) has one or more. The 0- and 1-lepton channels also make use of a dedicated $t\bar{t}$ control region, described in Section 7.1.4. A complete overview of the different analysis regions is given in Table 7.1.

Channel	Analysis Regions					
	$250 < p_T^V < 400 \text{ GeV}$		$p_T^V \geq 400 \text{ GeV}$			
	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 [124]. 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 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.

7.1.1 Data & Simulated Samples

The analysis uses pp collision data recorded between 2015 and 2018 by the ATLAS detector [32] during Run 2 at the LHC. This dataset corresponds to an integrated luminosity of 139 fb^{-1} .

Data from centre-of-mass energy $\sqrt{s} = 13 \text{ TeV}$ proton-proton collisions at the LHC recorded over the course of Run 2 were used for the analysis. The resulting dataset corresponds to a total integrated luminosity of 139 fb^{-1} (see Fig. 3.4).

An overview of the MC simulated samples used in the analysis is given in Table 7.2. These samples are used to model the signal and background processes relevant to the

analysis, with the exception of the multijet background which is modelled using a data-driven technique. Data and simulated events are reconstructed using the same algorithms, and a reweighting is applied to the simulated events in order to match the pile-up distribution observed in the data.

7.1.2 Object Reconstruction

The presence of neutrinos in the $WH \rightarrow \ell\nu b\bar{b}$ and $ZH \rightarrow \ell\ell b\bar{b}$ signatures can be inferred from a momentum imbalance in the transverse plane Section 3.4.5. The vector boson transverse momentum p_T^V is reconstructed as the missing transverse energy E_T^{miss} in the 0-lepton channel, as the magnitude of the summed $\mathbf{E}_T^{\text{miss}}$ and charged-lepton momentum in the 1-lepton channel, and as the transverse momentum of the 2-lepton system in the 2-lepton channel (see Section 3.4.5).

Leptons are used for the channel classification and to select relevant events as outlined in Section 7.1.3. Electrons and muons are reconstructed as outlined in Section 3.4.4. Electron identification follows the approach outlined in Ref. [108]. In addition to the likelihood-based method described in Section 3.4.4, *baseline* electrons are required to satisfy $p_T > 7 \text{ GeV}$, $|\eta| < 2.47$, $s(d_0) < 5$, and $|z_0 \sin(\theta)| < 0.5 \text{ mm}$. *Signal* electron additionally are required to satisfy a tighter likelihood identification selection. Muons are required to satisfy $p_T > 7 \text{ GeV}$, $|\eta| < 2.7$, $s(d_0) < 3$, and $|z_0 \sin(\theta)| < 0.5 \text{ mm}$. *Baseline* muons are required to pass the ‘loose’ identification described in Ref. [75], while *signal* muons are required to pass the ‘medium’ identification working point. All signal leptons are required to additionally satisfy a $p_T > 27 \text{ GeV}$ selection criteria, except for muons in the 1-lepton channel where a cut of 25 GeV is used. The number of baseline leptons is used to categorise the event into the 0-, 1- or 2-lepton channels. The 1- and 2-lepton channels additionally require one signal lepton to be present.

The track-jets matched to the Higgs candidate are b -tagged using the MV2c10 b -tagging algorithm [61, 77, 161]. MV2c10 is a machine learning algorithm using a Boosted Decision Tree (BDT) which is tuned to achieve an average b -jet efficiency of 70% on simulated $t\bar{t}$ events. At this efficiency working point, rejection factors for c -jets and light-jets are approximately 9 and 304 respectively. The MV2 algorithm takes inputs from the outputs of a number of low-level algorithms (IPxD, SV1 and JetFitter). The outputs of the low-level algorithms are provided as inputs to the boosted decision tree. The efficiency of the tagging algorithm is calibrated to events

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 [127] + GoSAM [129] + MnLO [130, 131]	NNPDF3.0NLO ^(*) [99]	Pythia 8.212 [101]	AZNLO	NNLO(QCD) + NLO(EW) [132–138]
$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$	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 [139–143]
Top quark ($m_t = 172.5$ GeV)					
$t\bar{t}$	Powheg-Box v2 [127, 144]	NNPDF3.0NLO	Pythia 8.230	A14 [102]	NNLO+NNLL [145]
s-channel	Powheg-Box v2 [127, 146]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [147]
t-channel	Powheg-Box v2 [127, 146]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [148]
Wt	Powheg-Box v2 [127, 149]	NNPDF3.0NLO	Pythia 8.230	A14	Approximate NNLO [150]
Vector boson + jets					
$W \rightarrow \ell\nu$	SHERPA 2.2.1 [151–154]	NNPDF3.0NNLO	SHERPA 2.2.1 [155, 156]	Default	NNLO [157]
$Z/\gamma^* \rightarrow \ell\ell$	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 7.2: Signal and background processes with the corresponding generators used for the nominal samples [124]. 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 [158] 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 [159, 160].

2043 in data [162–164]. The jet tagging strategy relies on extensive studies into track-jet
2044 b -tagging in boosted topologies [165, 166].

2045 The jet flavour labelling scheme is described in Section 3.4.3.

2046 7.1.3 Selection Criteria

2047 An extensive list of selection cuts are applied to each event in order to reject
2048 background events whilst retaining as many signal events as possible. A full list of
2049 selection cuts applied to the different analysis regions is given in Table 7.3, while
2050 some key selections are listed below.

2051 All channels are require events with at least one large- R jet with $p_T > 250 \text{ GeV}$
2052 and $|\eta| < 2.0$. The vector boson transverse momentum is also required to satisfy
2053 $p_T^V > 250 \text{ GeV}$. The Higgs candidate is chosen as the highest p_T large- R jet satisfying
2054 these requirements. As mentioned, the candidate large- R jet is required to have
2055 two ghost-assciated and b -tagged variable-radius track-jets. These track-jets are
2056 required to have at least two constituent tracks with $p_T > 500 \text{ MeV}$ and $|\eta| < 2.5$.
2057 The track-jets themselves must satisfy $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$.

2058 In the 0-lepton channel, trigger selections are applied using an E_T^{miss} trigger with a
2059 luminosity-dependent threshold between 70–110 GeV. In the 1-lepton electron sub-
2060 channel a combination of single electron triggers is used with minimum p_T thresholds
2061 between 24–26 GeV. In the muon sub-channel the same E_T^{miss} trigger as the 0-lepton
2062 channel is used. Since muons are not used for the E_T^{miss} trigger calculations, this
2063 is in effect a p_T requirement on the muon-neutrino system, which in the analysis
2064 phase space is more efficient than a single-muon trigger. The 2-lepton channel uses
2065 the same triggering strategy as the 1-lepton channel. In all channels, the trigger
2066 selections applied are fully efficient for events selected using the full requirements in
2067 Table 7.3.

2068 The combined selections in Table 7.3 result in a signal efficiency ranging from 6–16%
2069 for the WH and ZH processes depending on the channel and p_T^V bin.

Selection	0 lepton channel	1 lepton channel	2 leptons channel
Trigger	E_T^{miss}	e sub-channel Single electron	μ sub-channel E_T^{miss}
Leptons	0 <i>baseline</i> leptons	$p_T > 27 \text{ GeV}$ 1 <i>signal</i> lepton no second <i>baseline</i> lepton	2 <i>baseline</i> leptons among which $p_T > 25 \text{ GeV}$ both leptons of the same flavour - opposite sign muons
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	
b -tagged jets		leading two track-jets matched to the leading large- R must be b -tagged (MV2c10, 70%)	
m_1			$> 50 \text{ GeV}$
$\min[\Delta\phi(E_T^{\text{miss}}, \text{small-}R \text{ jets})]$	$> 30^\circ$		-
$\Delta\phi(E_T^{\text{miss}}, H_{\text{cand}})$	$> 120^\circ$		-
$\Delta\phi(E_T^{\text{miss}}, E_{T, \text{trk}}^{\text{miss}})$	$< 90^\circ$		-
$\Delta y(V, H_{\text{cand}})$	-		$ \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 [124]. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region.

2070 7.1.4 Control Regions

2071 The $t\bar{t}$ process presents a major background in the 0- and 1-lepton channels. In these
 2072 events, the Higgs candidate is often reconstructed from a correctly tagged b -jet from
 2073 the top decay $t \rightarrow Wb$, and an incorrectly tagged c - or light-jet from the subsequent
 2074 decay of the W , as shown in Fig. 7.1.

not sure
where the
0L ttbar
ETmiss
comes from

2075 The only known decay mode of the top quark is via the weak force to a W and
 2076 a down-type quark. (it is the only quark heavy enough to decay into an on-shell
 2077 W). Overwhelmingly (96% of the time) the down-type quark is a b -quark Hence, the
 2078 second top quark is also likely to result in a second tagged b -tagged track-jet outside
 2079 of the large- R Higgs candidate. To ensure sufficient $t\bar{t}$ rejections, 0- and 1-lepton
 2080 channel signal regions are defined using a veto on events with b -tagged track-jets
 2081 outside the Higgs-jet candidate. These events are used to construct a control region
 2082 (CR) which is enriched in $t\bar{t}$ events. The CR is used to constrain the normalisation
 2083 of the $t\bar{t}$ background in the fit.

2084 7.1.5 Background Composition

2085 After the selections described in Section 7.1.3 the number of background events
 2086 mimicking the VH , $H \rightarrow b\bar{b}$ signal is greatly reduced. However, the number of
 2087 background events still greatly outnumbers that of signal events. The background
 2088 processes are channel dependent. In the 0-lepton channel the dominant sources of
 2089 backgrounds are $Z+jets$ ($Z \rightarrow \nu\nu$) and $t\bar{t}$, with $W+jets$ and diboson events being
 2090 subdominant. In the event of $W \rightarrow \tau\nu$, and subsequent hadronic decay of the τ or
 2091 lack of successful reconstruction/selection of the leptonic decay products, $W+jets$
 2092 can also contribute to the 0-lepton channel. $t\bar{t}$ and $W+jets$ (with a leptonic decay
 2093 of the W as in $W \rightarrow \ell\nu$) are dominant in the 1-lepton channel, while single-top is
 2094 subdominant. In the 2-lepton channel, $Z+jets$ ($Z \rightarrow \ell\ell$) is again dominant followed
 2095 by $Z Z$ diboson events.

2096 The diboson background VV consists primarily of WZ and ZZ events in which the
 2097 Z decays to a pair of b -quarks. This process very closely matches the signal, with
 2098 a resonant peak occurring at $m_Z = 91$ GeV and so is considered as an irreducible
 2099 background ($V+b$ -jets is also irreducible).

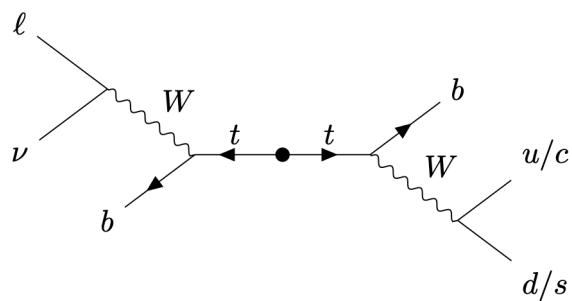
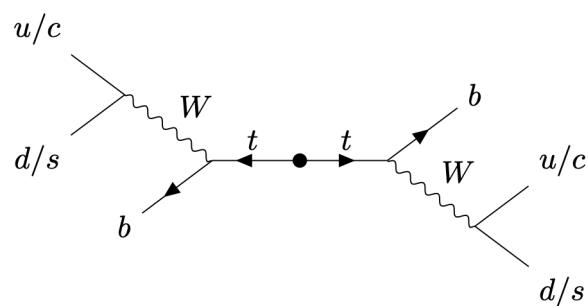
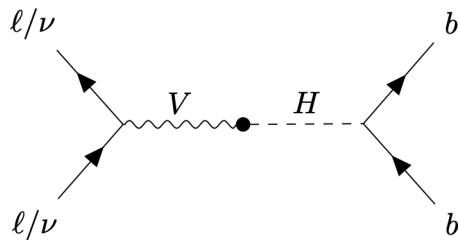


Figure 7.1: Diagrams of the signal process (top) and the 0-lepton and 1-lepton $t\bar{t}$ backgrounds (middle, bottom). Objects to the right of centre are reconstructed within the large- R jet. For the backgrounds, the large- R jet contains a mis-tagged c - or light-jet.

2100 The $t\bar{t}V$, $t\bar{t}H$ and multijet backgrounds are negligible in the analysis phase space
2101 after the selections have been applied, with the exception of the 1-lepton electron
2102 sub-channel, in which multijet background is not ignored. The multijet background
2103 is made up of jets with semileptonic heavy-flavour-hadron decays (e.g. $b \rightarrow c\ell\nu$) and
2104 jets which are mis-tagged by the flavour tagging algorithm MV2c10.

2105 The contributions from the different backgrounds are modelled using Monte Carlo
2106 event generators and the impacts on the analysis are studied in Section 7.2. The
2107 multijet background is not modelled but instead estimated using a data-driven
2108 technique.

2109 7.2 Systematic Uncertainties & Background 2110 Modelling

2111 Systemic uncertainties are extensively employed to give the fit model described in
2112 Section 7.3 enough flexibility to account for inaccuracies in the various inputs to the fit.
2113 Two main types of systematic uncertainty are considered: experimental and modelling.
2114 Experimental uncertainties arise due to the limited due to limited detector precision,
2115 imperfect reconstruction algorithms (in particular the b -tagging algorithms), and
2116 due to the imperfect measurement of pile-up and integrated luminosity. Modelling
2117 is the simulation processes relevant to the analysis using Monte-Carlo (MC) event
2118 generators, and is used to predict the outcome of the analysis. Modelling uncertainties
2119 arise due to the imperfections in the simulation of signal and background events, for
2120 example differences between event generators, or use of different model parameters
2121 when producing simulated events. In order to observe a certain process, for example
2122 VH , $H \rightarrow b\bar{b}$, an increase in the number of observed events with respect to the
2123 background-only hypothesis is looked for. The excess is often relatively small against
2124 the total number of background events, and hence accurate modelling of the expected
2125 number of background and signal events is crucial for successfully performing the
2126 analysis. Particular care is paid to the uncertainties on the modelling predictions as
2127 discussed in this section.

2128 Modelling uncertainties are described in detail in the following sections. Modelling
2129 uncertainties: *Nominal* samples are used as a reference to which different vari-

2130 ations can be compared. The nominal samples are chosen as the best possible
2131 representation of the underlying physical process. *Alternative* samples are used to
2132 understand inaccuracies that may be present in the nominal samples. Some aspect
2133 of the nominal model is varied, and the discrepancy with respect to the nominal
2134 model is quantified. The discrepancy is used to systematic uncertainty associated
2135 with the model parameter which was changed.

2136 Modelling studies involving c - and light-jets is hampered by the low available statistics
2137 of jets pass the analysis selections, due to the high rejection rates of the b -tagging
2138 algorithm MV2c10. For modelling studies therefore, truth tagging (TT) is employed
2139 to ensure sufficient numbers of jets are available to calculate uncertainties. TT works
2140 by computing a 2-dimensional efficiency map using the jet p_T and jet η . The two
2141 leading track-jets associated to the large- R jet automatically passes the b -tagging
2142 requirement, and are weighted based on their p_T and η using the pre-calculated
2143 efficiency map.

2144 7.2.1 Sources of Systematic Uncertainties

2145 This section briefly describes the different sources of uncertainty in the predictive
2146 model used in the analysis, and how each source of uncertainty is implemented within
2147 the analysis framework. Considered sources of systematic uncertainty are listed in
2148 Table 7.4. For each source of uncertainty, acceptance and shape uncertainties are
2149 derived.

2150 QCD Scales

2151 The $V+jets$ matrix element calculations contains infrared and ultraviolet divergences.
2152 These are handled by introducing spurious parameters corresponding to the renormalisation scale (μ_R) and factorisation scale (μ_F). Physical observables are not
2153 dependent on these parameters when using the infinite perturbation series expansion,
2154 however at some fixed order in QCD a limited dependence is present. To assess the
2155 impact of this, both μ_R and μ_F are independently varied from their nominal values
2156 by factors of 0.5 and 2 to account for higher order corrections to the calculation of
2157 the matrix element used to simulate the process.

2159 PDF Sets

2160 Parton distribution functions (PDFs) specify the probability of finding a parton
2161 with a given momentum inside a hadron (in this case, inside colliding protons).
2162 PDFs have to be derived from data and are a significant source of uncertainty in
2163 analyses of hadronic collision data. There are three sources of PDF uncertainties:
2164 the statistical and systematic errors on the underlying data used to derive the PDFs,
2165 the theory which is used to describe them (which is based on some fixed order
2166 perturbative QCD expansion), and finally the procedure which is used to extract the
2167 PDFs from the data. PDF-related uncertainties were derived following Ref. [158].
2168 This involves considering 100 PDF replicas which, when combined, form a central
2169 value and associated uncertainty, and also in parallel direct changes to the central
2170 values of PDFs using the MMHT2014 [167] and CT14NLO [168] PDF sets.

2171 Event Generator

2172 The choice of parton shower (PS) and underlying event (UE) generators can affect
2173 the analysis outcome. Changing these models modifies several aspects of the event
2174 generation at the same time, such as the accuracy of matrix element predictions and
2175 different approaches to parton showering. This change tends to lead to the largest
2176 discrepancy with respect to the nominal samples.

2177 Resummation and Merging Scales

2178 Resummation is a technique used in QCD to help cope with calculations involving
2179 disparate energy scales, and involves the introduction of an associated resummation
2180 scale, the choice of which introduces some systematic uncertainty into the model.
2181 Parton showering models are accurate when simulating low- p_T radiation, however
2182 inaccuracies start to arrive when simulating hard emissions. To combat this, par-
2183 ton showering models utilise more precise matrix element calculations above some
2184 momentum threshold. The choice of threshold, or *merging scale* introduces some
2185 uncertainty into the final result. Resummation (QSF) and merging (CKKW) scale
2186 variations are available for a subset of the SHERPA samples. The number of available
2187 events is significantly lower than the number of events in the nominal sample, and no

2188 statistically significant discrepancy with respect to the nominal samples is observed.
2189 The corresponding uncertainties and therefore neglected.

2190 7.2.2 Implementation of Variations

2191 Modelling variations are implemented in different ways, depending on the associated
2192 uncertainty. Table 7.4 lists the different sources of uncertainty described in Section 7.2.1
and for each lists the implementation. As production of high-stastic MC samples

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 7.4: 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.

2193

2194 is computationally expensive, a technique in state of the art simulation packages
2195 is to store some sources of variation as internal weights, which can be generated
2196 alongside the nominal samples, saving computation time. The nominal sample then
2197 effectively contains information about an ensemble of different samples, corresponding
2198 to different model parameters, which are accessible via reweightings. When filling
2199 histograms for the variations, bins are incremented by the internal weight of the
2200 event associated with the variation in question.

2201 While the inclusion of internal weight variation in MC event generators has decreased
2202 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently
2203 some sources of systematic uncertainty that are unable to be stored as internal weight
2204 variations due to technical limitations. Two examples are the choice of resummation
2205 and merging scales. A method to parameterise the systematic variation using
2206 one sample, and to then apply this parameterisation to another sample, has been

2207 developed by ATLAS [169]. This method was used to derive resummation and
2208 merging uncertainties for the nominal SHERPA 2.2.1 sample, using a previous (lower
2209 statistic) SHERPA 2.1 alternative sample. The resulting uncertainties were studied
2210 and found to be negligible in comparison with systematics from other sources.

2211 7.2.3 Vector Boson + Jets Modelling

2212 After event selection, the $V+jets$ background is a dominant background in all three
2213 analysis channels as described in Section 7.1.5. The $V+jets$ samples are split into
2214 categories depending on the truth flavour of the track-jets which are ghost-associated
2215 to the large- R jet Higgs candidate. The categories are $V+bb$, $V+bc$, $V+bl$, $V+cc$,
2216 $V+cl$, $V+ll$, and $V+hf$ refers collectively to the categories containing at least one
2217 b - or c -jet. $V+bb$ is dominant generally accounting for 80% of the jets, while $V+hf$
2218 accounts for around 90% of jets. The full flavour composition breakdown for each
2219 channel and analysis region are given in Tables 7.5, 7.7 and 7.8.

2220 In order to access uncertainties associated with the use of MC generators, variations
2221 of the data are produced using alternative generators or variation of nominal
2222 generator parameters as described in Section 7.2.2. As described in Section 7.1.1,
2223 the nominal MC event generator used for $V+jets$ events is SHERPA 2.2.1, while
2224 MADGRAPH5_AMC@NLO+PYTHIA8 (which uses a different parton showering
2225 model) is used as an alternative generator.

2226 Modelling systematics can have several impacts, including affecting the overall
2227 normalisation for different processes, and the relative acceptances between different
2228 analysis regions (i.e. migrations between HP and LP SRs, between the SR and CR,
2229 and between p_T^V bins), and the shapes of the m_J distributions. Since the fit model
2230 fits only the large- R jet mass m_J to data, all shape uncertainties are estimated with
2231 respect to this observable. Several sources of uncertainty, summarised in Section 7.2.1,
2232 have been assessed.

Sample	$M_{p_T}^V$ HP SR	$H_{p_T}^V$ HP SR	$M_{p_T}^V$ LP SR	$H_{p_T}^V$ LP SR	$M_{p_T}^V$ CR	$H_{p_T}^V$ 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 7.5: 0-lepton $W+jets$ nominal sample flavour composition and total event yield [170].

Sample	$M_{p_T}^V$ HP SR	$H_{p_T}^V$ HP SR	$M_{p_T}^V$ LP SR	$H_{p_T}^V$ LP SR	$M_{p_T}^V$ CR	$H_{p_T}^V$ 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 7.6: 1-lepton $W+jets$ nominal sample flavour composition and total event yield [170].

Channel	$M_{p_T}^V$ HP SR	$H_{p_T}^V$ HP SR	$M_{p_T}^V$ LP SR	$H_{p_T}^V$ LP SR	$M_{p_T}^V$ CR	$H_{p_T}^V$ 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 7.7: 0-lepton $Z+jets$ nominal sample flavour composition and total event yield [170].

Channel	$M_{p_T}^V$	$H_{p_T}^V$	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 7.8: 2-lepton $Z+jets$ nominal sample flavour composition and total event yield [170].

2233 **Acceptance Uncertainties**

2234 Several different types of acceptance uncertainties have been calculated and imple-
 2235 mented as nuisance parameters in the fit. These account for uncertainty in the overall
 2236 number of events in each channel, and for the migration of events between different
 2237 analysis regions. The acceptance uncertainties relevant to the $V+jets$ processes are
 2238 summarised below.

2239 • **Overall normalisation:** only relevant where normalisation cannot be left
 2240 floating (determined as part of the fit). The $V+hf$ component is left floating in
 2241 the fit. For other components, independent normalisations used for $W+hf$ and
 2242 $Z+hf$. The contributions are mainly determined by the 1-lepton (for $W+hf$)
 2243 and 2-lepton (for $Z+hf$) SRs respectively and then extrapolated to 0-lepton
 2244 channel.

2245 • **SR-to-CR relative acceptance:** the uncertainty on the normalisation of the
 2246 signal region due to events migrating between the signal and control regions.

2247 • **HP-to-LP relative acceptance:** the uncertainty on the normalisation of the
 2248 high-purity (HP) signal region due to events migrating between the high- and
 2249 low-purity signal regions.

2250 • **Medium-to-high p_T^V relative acceptance:** describes any shape effect in p_T^V
 2251 distribution, given that the analysis only uses two p_T^V bins (medium and high).

2252 • **Flavour relative acceptance:** for each flavour $V+xx$, where $xx \in \{bc, bl, cc\}$
 2253 the ratio of $V+xx/V+bb$ events is calculated. This corresponds to the uncer-
 2254 tainty of Vbb events due to the miss-tagging of other flavours Vxx .

2255 • **Channel relative acceptance:** corresponding to the uncertainty in the nor-
 2256 malisation of $V+jets$ events events due to the migration of events between
 2257 channels.

2258 The uncertainties arising from the different sources described in Section 7.2.1 are
 2259 summed in quadrature to give a total uncertainty on each region. A summary of the
 2260 different acceptance uncertainties that were derived in this way and subsequently
 2261 applied in the fit are given in Table 7.9. An effort has been made, wherever possible,
 2262 to harmonise similar uncertainties across different analysis regions and channels.

V+jets Acceptance Uncertainties				
Boson	W		Z	
Channel	0L	1L	0L	2L
Vbb Norm.	30%	-	-	-
SR/CR	90% [†]	40% [†]	40%	-
HP/LP	18%		18%	-
High/Medium p_T^V	30%	10%*	10%	
Channel Extrap.	20%	-	16%	-
Vbc/Vbb	30%			
Vbl/Vbb	30%			
Vcc/Vbb	20%			
Vcl Norm.	30%			
Vl Norm.	30%			

Table 7.9: V+jets acceptance uncertainties [170]. W+jets SR and CR uncertainties marked with a superscript † are correlated. The 1L W+jets H/M uncertainty marked by * is applied as independent and uncorrelated NPs in both HP and LP signal regions. The 0L W+jets Wbb Norm uncertainty is only applied when a floating normalisation for Wbb cannot be obtained from the 1L channel. A 30% uncertainty for $Z \rightarrow b\bar{b}$ norm is applied in the 1L channel when a floating normalisation for $Z \rightarrow b\bar{b}$ cannot be obtained from the 0L or 2L channels.

2263 Shape Uncertainties

2264 In order to derive shape uncertainties (which as the name suggests affect shapes but
 2265 not overall normalisations of distributions), the following procedure is carried out.
 2266 Normalised distributions of the reconstructed large- R Higgs candidate jet mass m_J
 2267 are compared for the nominal sample and variations. For each variation, the ratio of
 2268 the variation to nominal is calculated, the up and down variations are symmetrised,
 2269 and an analytic function is fit to the symmetrised ratio. If different analysis regions
 2270 or channels show the same pattern of variation, a common uncertainty is assigned.
 2271 An example of a significant source of uncertainty, arising from choice of factorisation
 2272 scale μ_R is shown in Fig. 7.2. HP SRs split into medium and high p_T^V bins are shown
 2273 for the 0-lepton channel for $W+\text{hf}$ and $Z+\text{hf}$ jets. The 0- and 1-lepton channels for
 2274 the $W+\text{hf}$ contribution and the 0- and 2-lepton channels for the $Z+\text{jets}$ contribution
 2275 are merged, since the shapes in m_J are consistent across channels. An exponential
 2276 function $e^{p_0 + p_1 x} + p_2$ has been fitted to the ratio of the normalised distributions.
 2277 The magnitude of the variation does show p_T^V dependence, and so two separate
 2278 uncertainties are added in the fit, and applied individually in each p_T^V region.

2279 The shape uncertainties for μ_R were derived on the SRs but are also applied to the
 2280 CRs, as the low statistics in the CRs make it difficult to derive dedicated shape
 2281 uncertainties. All the shape uncertainties are fully correlated accross regions.

2282 A comparison of the m_J shapes between SHERPA and MADGRAPH is shown in
 2283 Fig. 7.3. The plots are split by process and channel, but merged in SR purity and p_T^V
 2284 bins reflecting similarities between the m_J shapes across these regions. Due to the low
 2285 statistics available for the alternate MADGRAPH sample, and the lack of statistically
 2286 significant variation between the samples, no associated shape uncertainty is added
 2287 to the fit in this case.

2288 The impacts of variations in the factorisation scale μ_F and the choice of PDF set on
 2289 m_J shape were also found to be negligible in comparison with μ_R and are hence not
 2290 associated uncertainty was added to the fit.

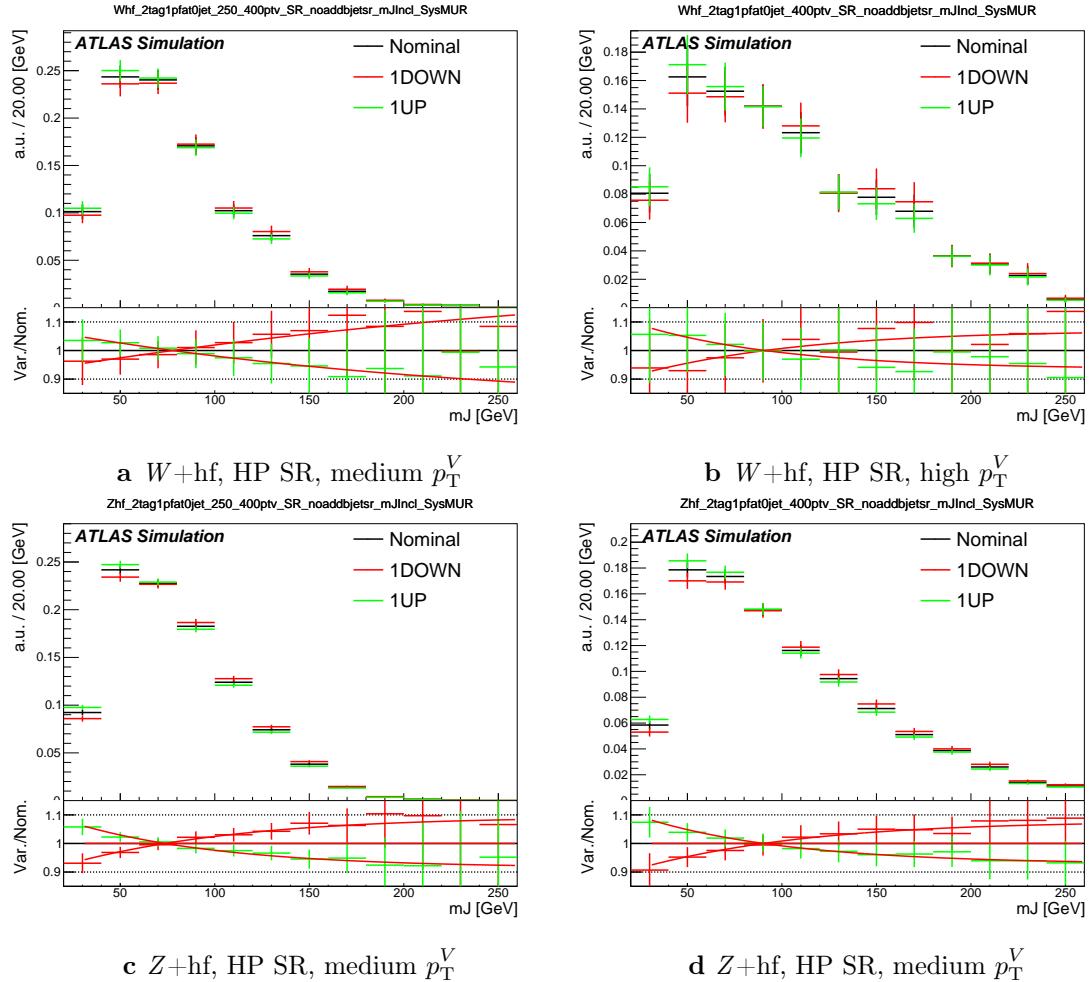


Figure 7.2: Normalised leading large- R jet mass distribution from Z and $W+hf$ processes in the HP SR of the 0-lepton channel [170]. The renormalisation scale μ_r has been varied by a factor of 2 (1up) and 0.5 (1down). An exponential function is fitted to the ratio between the nominal and variation samples.

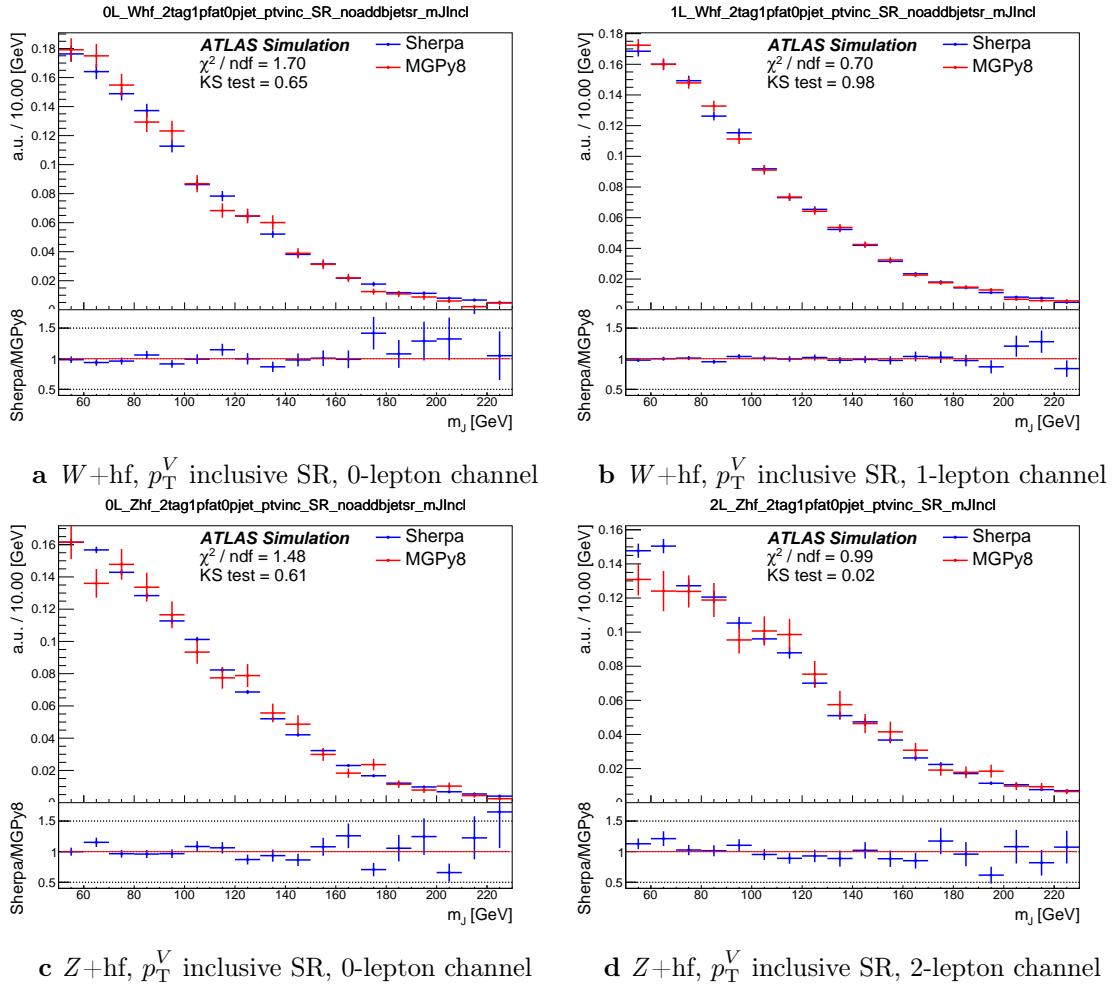


Figure 7.3: The comparison on m_J shapes between SHERPA and MADGRAPH samples from $V+hf$ process in p_T^V inclusive signal regions [170]. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.

2291 7.2.4 Diboson Modelling

2292 The uncertainties for the diboson background generally follows that of $V+jets$. How-
2293 ever an alternative sample was generated using POWHEG interfaced with PYTHIA8,
2294 using the AZNLO shower tune with the CTEQ6L1 PDFs [171]. Unlike SHERPA,
2295 POWHEG models the off-shell Z contribution at NLO.

2296 Acceptance and shape uncertainties are derived in an analogous fashion to $V+jets$
2297 as described below.

2298 Acceptance Uncertainties

2299 Diboson acceptance uncertainties are summarised in Table 7.10. Variations from μ_R ,
2300 μ_F , PDF choice and alternative generator are considered and are combined combined
2301 through a sum in quadrature as described in Section 7.2.3. The largest modification
2302 to the nominal acceptance results from the POWHEG+PYTHIA8 alternate sample,
2303 which modifies several parts of the generative model at the same time. Since the
2304 diboson contribution to the $t\bar{t}$ control region is small, no SR-to-CR relative acceptance
2305 uncertainty is required.

2306 For the WZ contribution, uncertainties are derived using the 1-lepton channel and
2307 applied to all three channels. An additional 8% channel migration uncertainty
2308 is applied on the 0-lepton channel. For the ZZ contribution, the normalisation
2309 uncertainty is calculated using the 2-lepton channel and applied to all three channels.
2310 The 0- and 1-lepton channels have a similar HP-to-LP relative acceptance uncertainty
2311 of 18%. The 1-lepton medium-to-high p_T^V relative acceptance is based off the value
2312 obtained from the 2-lepton channel. 30% and 18% channel migration uncertainties
2313 are applied in the 0- and 1-lepton channels respectively.

2314 Since the contribution from WW is small, dedicated studies are not performed, but
2315 a 25% normalisation uncertainty is applied in all the three channels which is based
2316 on the modelling studies performed for the previous analysis [108].

Diboson Acceptance Uncertainties						
Bosons	WZ			ZZ		
Channel	0L	1L	2L	0L	1L	2L
Normalisation	16%			10%		
HP/LP	18%			18%		
High/Medium	10%			6%	18%	
Channel Extrap.	8%	-	-	30%	18%	-

Table 7.10: Diboson acceptance uncertainties [170]. All uncertainties except channel extrapolation uncertainties are fully correlated between ZZ and WZ processes and channels.

2317 Shape Uncertainties

2318 Diboson shape uncertainties are derived in a similar fashion to $V+jets$. Only the
2319 uncertainties associated with systematic variation of μ_R and the event generator
2320 have a non-negligible impact on the m_J shape. Variation of μ_R produces consistent
2321 m_J shape impacts across all regions and channels, and hence only a single associated
2322 uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted to the
2323 symmetrised ratio.

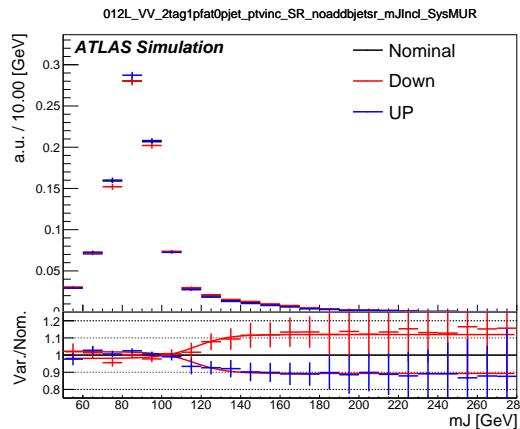


Figure 7.4: Normalised leading large- R jet mass distribution from WZ and ZZ process, merged among all the signal regions and lepton channels [170]. The renormalisation scale μ_R has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function.

2324 The comparison between the nominal SHERPA and alternative POWHEG+PYTHIA
2325 8 samples is shown in Fig. 7.5 for the 0- and 1-lepton channels for both WZ and

2326 ZZ processes. For these channels, the shape of m_J varies in opposite directions
2327 in the LP and HP signal regions. Shapes are similar between p_T^V bins, the 0- and
2328 1-lepton channels and for WZ and ZZ . A third order polynomial is fitted to the
2329 ratio, and this function transitions to a constant piecewise function in the high
2330 mass region to accurately represent the shape taking into account large statistical
2331 uncertainties. Dependence on event generator was found to be negligible within
2332 statistical uncertainty in the 2-lepton channel. All diboson shape uncertainties are
2333 fully correlated in the fit.

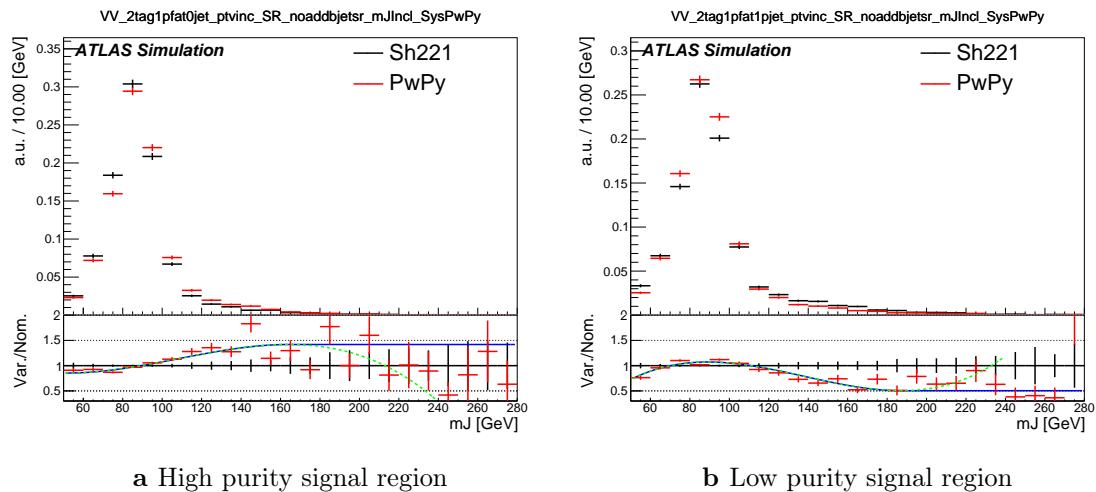


Figure 7.5: The comparison on m_J shapes between SHERPA and POWHEG+PYTHIA 8 samples from WZ and ZZ process in high and low purity signal regions [170]. p_T^V regions and 0- and 1-lepton channels are merged. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after a protection is added in the high mass region.

2334 7.3 Statistical Treatment

2335 Selected events are used to perform a statistical test of the background-only hy-
2336 pothesis, i.e. a model which does not include the VH , $H \rightarrow b\bar{b}$ process. The test
2337 involves a binned global maximum-profile-likelihood fit from the model to the data
2338 using the m_J distribution, and combines all the analysis regions defined in Table 7.1.
2339 The test is based on the profile likelihood ratio test statistic. The signal strength
2340 $\mu = \sigma/\sigma_{SM}$ is defined as the ratio between the observed and predicted cross-sections,
2341 where $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ corresponds

2342 to the SM prediction. It is a parameter of interest (POI) which acts to scale the
2343 total number of signal events.

2344 The present analysis makes use of two POIs. The first, μ_{VH}^{bb} , is the signal strength
2345 for the VH , $H \rightarrow b\bar{b}$ process, the primary process under investigation. The diboson
2346 production strength μ_{VZ}^{bb} for the VZ , $Z \rightarrow b\bar{b}$ process is measured simultaneously
2347 and provides a validation of the analysis apparatus used for the primary $H \rightarrow b\bar{b}$
2348 measurement. Alongside the two POIs, the predictive model depends on several
2349 uninteresting parameters which are not the primary target of measurement. These
2350 parameters are called nuisance parameters (NPs), collectively referred to as θ . Freely
2351 floating background normalisations are implemented as NPs and are also extracted
2352 during the fitting processes.

2353 7.3.1 Likelihood Function

2354 The statistical setup treats each bin as a Poisson counting experiment and is based on
2355 the ROOSTATS framework [172]. The combined likelihood over N bins is constructed
2356 as the product of Poisson probabilities in each bin. Considering the simplified case
2357 of a single signal strength parameter μ , and neglecting sources of systematic or
2358 statistical uncertainty, this is given by

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

2359 where s_i (b_i) is the expected number of signal (background) events in bin i , and n_i is
2360 the number of observed data events in bin i .

2361 Treatment of Uncertainties

2362 Systematic uncertainties can modify the predicted signal and background yields s_i
2363 and b_i . Each source of systematic uncertainty is taken into account by adding an
2364 additional NP θ_j to the likelihood in the form of a Gaussian cost function. The

2365 combined effect of the NPs is then

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

2366 where N_θ is the number of NPs, $\bar{\theta}_j$ is the nominal value of the j th NP, θ_j is the fitted
2367 value, and σ_j is the corresponding associated prior uncertainty on θ_j . As the fitted
2368 value of the θ_j deviates from its nominal value, a cost is introduced. The presence
2369 of NPs modifies the likelihood as

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

2370 The predicted signal and background yields are also modified by the presence of the
2371 NPs with

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

2372 For NPs which are left freely floating in the fit, no corresponding Gaussian constraint
2373 is added to the likelihood.

2374 Statistical uncertainty is also present, and implemented using a dedicated NP for
2375 each bin which can scale the background yield in that bin. Statistical NPs are also
2376 implemented using a Gaussian constraint.

2377 Smoothing and Pruning

2378 Systematic uncertainties are smoothed and pruned in the fit. Smoothing accounts
2379 for the large statistical uncertainty present in some bins that are subject to large
2380 fluctuations. The smoothing procedure relies on the assumption that the impact of
2381 systematics should be approximately monotonic and correlated between neighbouring
2382 bins.

2383 In addition to smoothing, pruning is the process of removing from the fit those
2384 systematics which only have a very small effect. This improves the stability of the
2385 fit by reducing the number of degrees of freedom. Acceptance uncertainties are
2386 pruned in a given region if they have a variation of less than 0.5%, or if the up and
2387 down variations have the same sign in that region. Shape uncertainties are pruned

2388 in a given region if the deviation in each bin is less than 0.5% in that region. In
2389 addition, acceptance and shape uncertainties are neglected in a given region for any
2390 background which makes up less than 2% of the total background in a given region.

2391 **Fit Procedure and Statistical Tests**

2392 The best-fit value of μ , denoted $\hat{\mu}$, is obtained via an unconditional maximisation
2393 of the likelihood. The likelihood is also used to construct a statistical test which can
2394 confirm or reject the background-only hypothesis. The test statistic q_μ is constructed
2395 from the profile likelihood ratio, as in

$$q_\mu = -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad (7.5)$$

2396 where $\hat{\mu}$ and $\hat{\theta}$ are chosen to maximise the likelihood \mathcal{L} , and the profile value $\hat{\theta}_\mu$ is
2397 obtained from a conditional maximisation fo the likelihood for a specific choice of
2398 $\mu = 0$ corresponding to the background-only hypothesis.

2399 The test statistic is used to construct a p -value which is used to confirm or accept
2400 the background-only hypothesis. The p -value is typically reported in terms of the
2401 significance Z , defined as the number of standard deviations for a Gaussian Normal
2402 distribution which will produce a one-sided tail integral equal to the p -value, as in

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 1 - \Phi(Z). \quad (7.6)$$

2403 Typically a value of $Z = 3$ constitutes *evidence* of a processes, while $Z = 5$ is required
2404 for a *discovery*. Alongside the p -value, the best-fit value of the signal strength $\hat{\mu}$ and
2405 its corresponding uncertainty are typically quoted, and compared to their expected
2406 values (see Section 7.3.3).

2407 **7.3.2 Background Normalisations**

2408 The normalisation of the largest backgrounds are left floating and are determined in
2409 the fit. The corresponding postfit background normalisations are listed in Table 7.11.
2410 A single normalisation factor is used for $W+\text{hf}$ and $Z+\text{hf}$, which constitue more

2411 than 90% of the total $V + \text{jets}$ background, since the use of independent factors in
2412 different channels were found to be compatible.

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

Table 7.11: Factors applied to the nominal normalisations of the $t\bar{t}$, $W + \text{hf}$, and $Z + \text{hf}$ backgrounds, as obtained from the likelihood fit [124]. The errors represent the combined statistical and systematic uncertainties.

2413 The normalisations and shapes of all other backgrounds, with the exception of the
2414 multijet background which is estimated using a data driven technique, are initialised
2415 using the simulated samples.

2416 7.3.3 Asimov Dataset & Expected Results

2417 The Asimov dataset is constructed by replacing the data with the sum of the signal
2418 and background predictions $n_i = s_i + b_i$. A fit to this dataset using the nominal
2419 values of the NPs from the simulation will recover the input values and is useful for
2420 studying constraints on and correlations between the NPs.

2421 Alternatively, a conditional fit to the Asimov dataset can be performed using values
2422 of the background NPs which are determined from an unconditional fit to data. The
2423 signal NPs and POIs are fixed at their nominal values from the SM simulation. The
2424 result of this fit can be used to calculate expected (median) significances, which can
2425 be compared to their observed values, as is done in Section 7.4.2.

2426 7.4 Results

2427 In the present analysis, the two signal strength parameters μ_{VH}^{bb} and μ_{VZ}^{bb} are extracted
2428 from a simultaneous maximisation of the likelihood described in Section 7.3. The
2429 results of the analysis are summarised in this section. Post-fit m_J distributions are

2430 shown in Section 7.4.1. The observed signal strengths are given in Section 7.4.2,
2431 along with observed and expected significances. Finally in Section 7.4.3 the impact
2432 of systematic uncertainties on the results is examined.

2433 7.4.1 Post-fit Distributions

2434 In addition to the observed significance and signal strength, it is also useful to study
2435 the post-fit m_J distributions to compare the simulation and data using the best-fit
2436 values $\hat{\mu}$ and $\hat{\theta}$. Post-fit m_J distributions are given for the signal regions in the 0-,
2437 1- and 2-lepton channels in Fig. 7.6. The LP and HP regions are merged for the
2438 0- and 1-lepton channels. The plots show large falling backgrounds, predominantly
2439 made up of W -jets and Z -jets events, and a signal distribution corresponding to
2440 the Standard Model Higgs boson peaking around $m_H = 125$ GeV. In general there is
2441 a good level of agreement between the simulation and data, indicating the fit model
2442 is performing as expected. Fig. 7.7 shows the post-fit plots for the $t\bar{t}$ control regions.
2443 Again, a good level of agreement is observed given the statistical uncertainties on
2444 the distributions.

2445 7.4.2 Observed Signal Strength & Significance

2446 The measured signal strength is computed as the ratio between the measured signal
2447 yield to the prediction from the SM. The combined result for all three lepton channels
2448 and all analysis regions is given for μ_{VH}^{bb} in Eq. (7.7), and for μ_{VZ}^{bb} is given in Eq. (7.8).
2449 Both results include a full breakdown of the systematic and statistical uncertainties.

$$\mu_{VH}^{bb} = 0.72^{+0.39}_{-0.36} = 0.72^{+0.29}_{-0.28}(\text{stat.})^{+0.26}_{-0.22}(\text{syst.}) \quad (7.7)$$

2450

$$\mu_{VZ}^{bb} = 0.91^{+0.29}_{-0.23} = 0.91 \pm 0.15(\text{stat.})^{+0.24}_{-0.17}(\text{syst.}) \quad (7.8)$$

2451 The results for μ_{VH}^{bb} and μ_{VZ}^{bb} agree with the expectation from the SM within their
2452 combined uncertainty. The μ_{VH}^{bb} measurement is dominated by statistical uncertainty,
2453 while the μ_{VZ}^{bb} measurement is dominated by systematic sources of uncertainty. These
2454 measured signal strength for μ_{VZ}^{bb} corresponds to an observed significance of 2.1
2455 standard deviations, with an expected (median) significance given the SM prediction

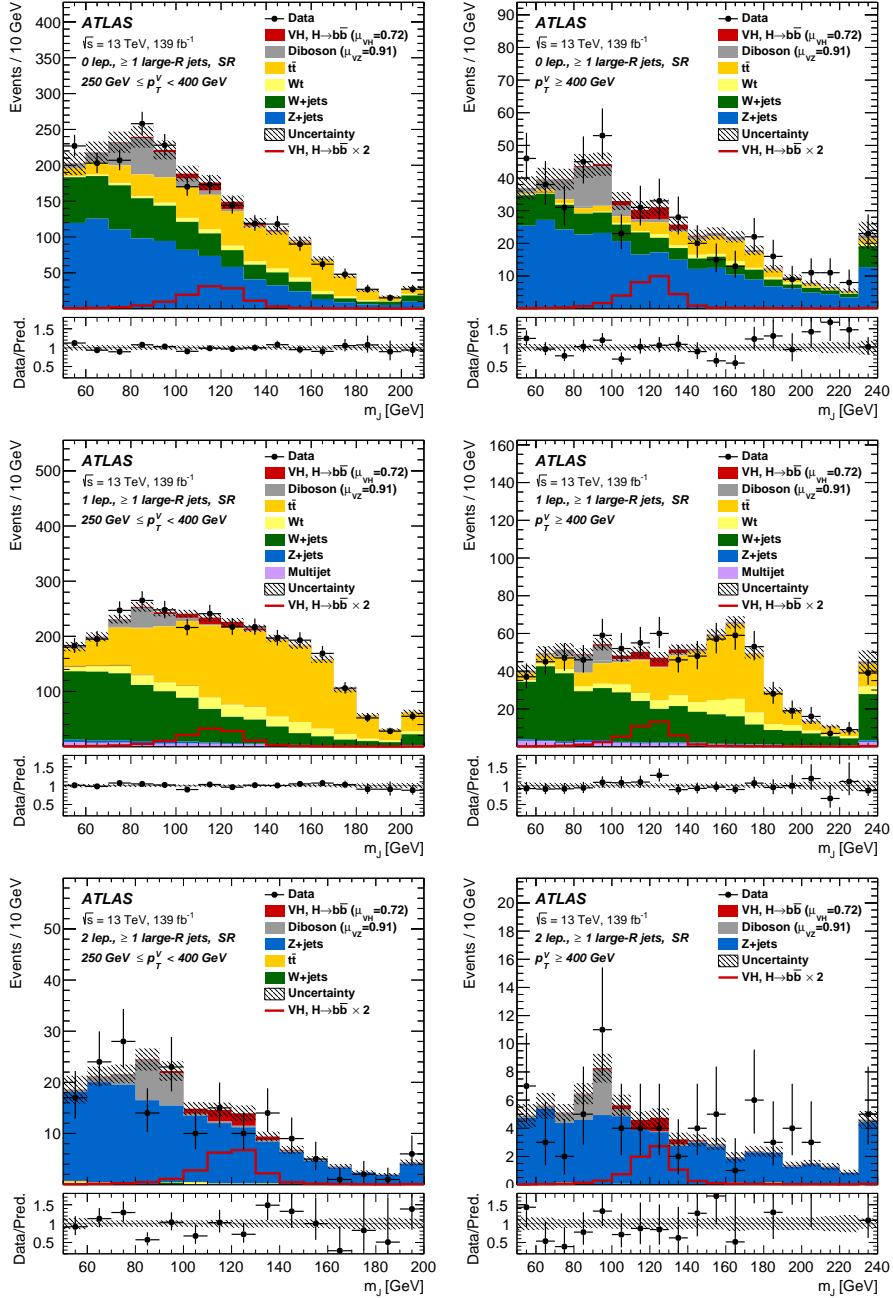


Figure 7.6: The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$ [124]. The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions 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.

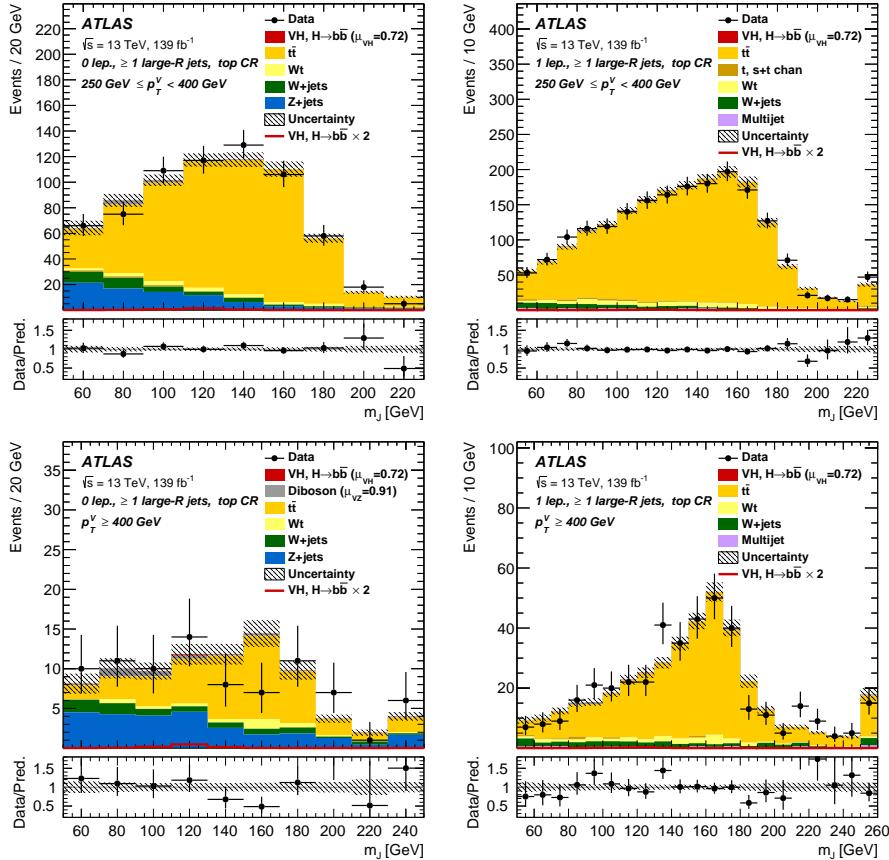


Figure 7.7: The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$ [124]. The background contributions 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.

of 2.7 standard deviations obtained using the method described in Section 7.3.3. The diboson observed (expected) signal strength significance is 5.4 (5.7). These results are summarised in Fig. 7.8, which shows the background-subtracted m_J distribution. A clear signal excess is visible around the Higgs mass of $m_H = 125$ GeV.

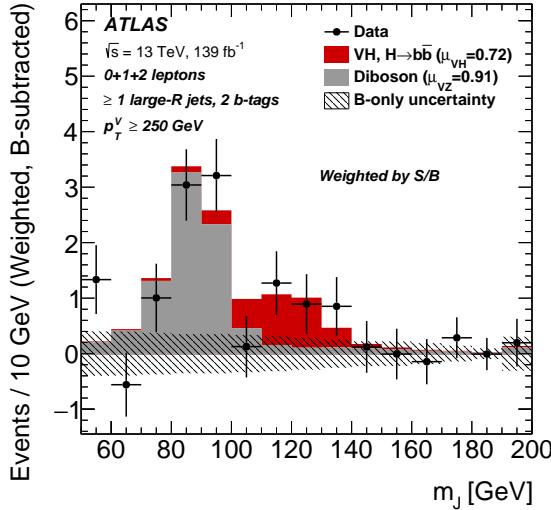


Figure 7.8: m_J distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes [124]. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected contribution of the associated WH and ZH production of a SM Higgs boson with $m_H = 125$ GeV is shown scaled by the measured combined signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and systematic uncertainty is indicated by the hatched band. This error band is computed from a full signal-plus-background fit including all the systematic uncertainties defined in Section 7.2, except for the VH/VZ experimental and theory uncertainties.

2460 Compatability Studies

Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting μ_{VH}^{bb} into three separate POIs, one for each channel. A simultaneous fit to the channel specific signal strengths can then be performed, which allows a comparison of the contributions from each channel. Fig. 7.9 compares the best-fit signal strengths. The 0- and 1-lepton channels show a signal strength which is consistent with the SM prediction, while the 2-lepton channel shows a small deviation within the 1σ

2467 uncertainty. Overall, good compatibility is observed via a χ^2 test with a corresponding
2468 p -value of 49%.

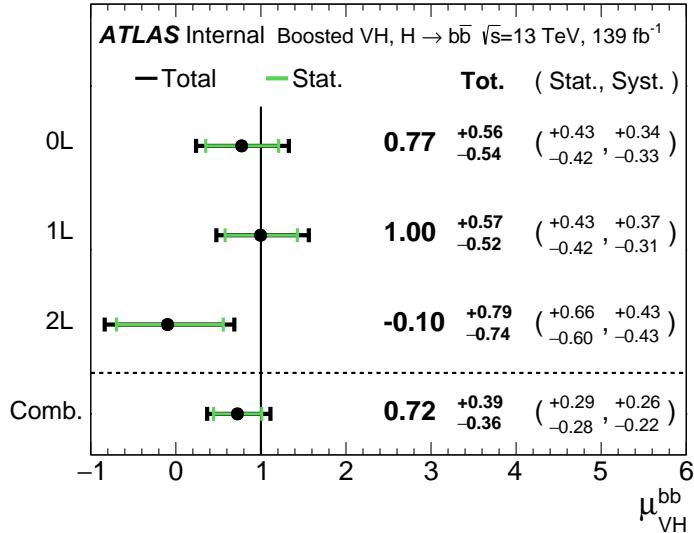


Figure 7.9: Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels splitted) and the default (1+1)-POI fit [124]. The compatibility of the three channels is evaluated via a χ^2 difference test and results in a p-value of 49%.

2469 7.4.3 Impact of Systematics

2470 The impact of systematic uncertainties on the final fitted value $\hat{\mu}_{VH}^{bb}$ is studied looking
2471 at the NP rankings, and the uncertainty breakdown.

2472 Fig. 7.10 shows the NP ranking, which is used to visualise which out of the many
2473 NPs involved in the fit have the largest impact on the sensitivity to the fitted POI.
2474 To obtain the ranking, a likelihood scan is performed for each NP θ_j . First, an
2475 unconditional fit is used to determine $\hat{\theta}_j$. From this best-fit point, the NP is varied
2476 in steps and the likelihood is recomputed until the $\pm 1\sigma_{\hat{\theta}_j}$ values are reached. For
2477 each corresponding value of θ_j , the change in the best-fit value of the POI, $\Delta\hat{\mu}_{VH}^{bb}$
2478 is calculated and used to rank the NPs. Also shown in Fig. 7.10 are the pulls and
2479 constraints for the highest ranked NPs.

2480 The experimental uncertainty on the signal large- R jet mass resolution (JMR) has
2481 the largest impact of any NP. It is a significant contributor to the overall uncertainty

2482 on μ_{VH}^{bb} in Eq. (7.7). JMR and jet energy scale (JES) uncertainties also have impacts
2483 for the $V+jets$ background and for the diboson background. The freely-floating
2484 $Z+hf$ normalisation is the second highest ranked NP, and is heavily constrained by
2485 the fit. The VZ POI μ_{VZ}^{bb} is also a significant NP when considering the primary μ_{VH}^{bb}
2486 measurement.

2487 The NP ranking highlights individual NPs which have a large impact on the POI
2488 measurement sensitivity. Complementary information is provided at a higher level
2489 by considering the overall impact of different groups of systematics. The groups
2490 are constructed from NPs which have similar physical origin. The impact on each
2491 group is calculated by running a fit with all the NPs in the given group fixed to their
2492 nominal values. The uncertainty on the POI extracted from this fit is subtracted
2493 in quadrature from the the uncertainty on the POI from the nominal fit, and the
2494 resulting values are provided as the impact for each group. The full breakdown
2495 for the observed impact of uncertainties on the μ_{VH}^{bb} signal strength is provided in
2496 Table 7.12. The total systematic impact is the difference in quadrature between the
2497 nominal uncertainty on μ_{VH}^{bb} and the combined statistical impact. The “data stat
2498 only” group fixes all NPs at their nominal value, while the total statistical impact
2499 fixes all NPs except floating normalisations. The floating normalisations group fixes
2500 only the NPs associated with normalisation which are left floating in the fit. The
2501 uncertainty on μ_{VH}^{bb} is dominated by combined statistcal effects (0.28), although the
2502 combined impact of systematics (0.24) is of a comparable size. The signal largest
2503 group is the data stat uncertainty (0.25), demonstrating that the analysis would
2504 benefit from an increased integrated luminosity or improved efficiency to select signal
2505 relevant events (recall from Section 7.1.3 the signal efficiency is in the range of
2506 10%). Of the experimental systematic sources of uncertainty, the dominant impact
2507 is the experimental uncertainties associated with the reconstruction of large- R jets
2508 (0.13). Other experimental sources of uncertainty are small in comparison. Modelling
2509 uncertainties also have a large contribution to the overall systematic uncertainty. The
2510 biggest contribution to the overall uncertainty is the combined statistical uncertainty
2511 on the simulated samples (0.09), which contain only a finite number of events. Out of
2512 the backgrounds, the $W+jets$ and $Z+jets$ have the highest (0.06) and second-highest
2513 (0.05) impact respectively.

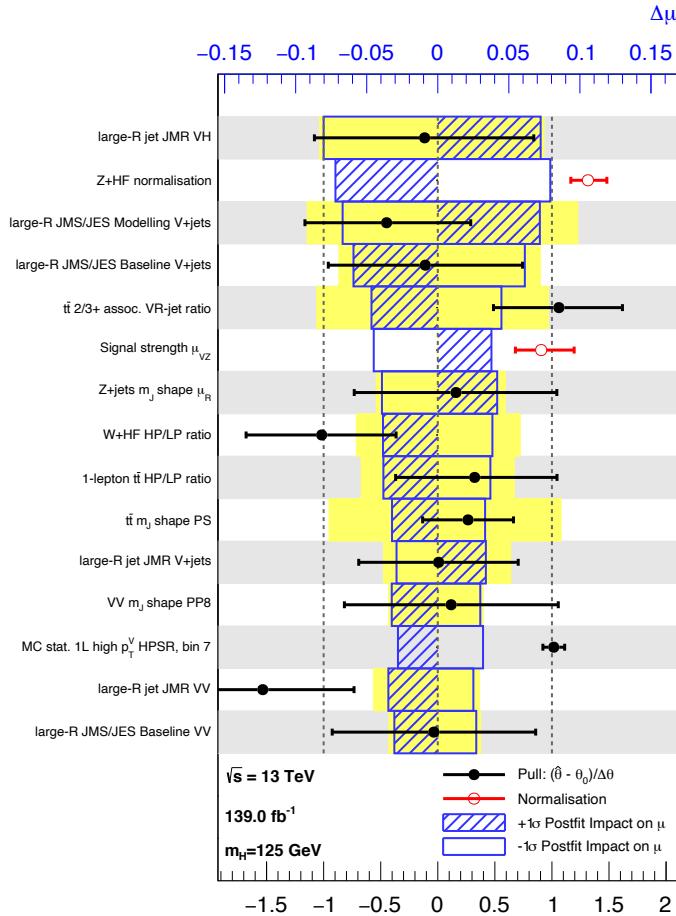


Figure 7.10: Impact of systematic uncertainties on the fitted VH signal-strength parameter μ_{VH}^{bb} sorted in decreasing order [170]. The boxes show the variations of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual nuisance parameter to its post-fit value modified 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.

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
<i>b</i> -tagging	<i>b</i> -jets	+0.028 / -0.004
	<i>c</i> -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
↔ $Z + \text{jets}$	+0.049 / -0.047	0.048
↔ $W + \text{jets}$	+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 7.12: Breakdown of the observed absolute contributions to the uncertainty on the signal strength μ_{VH}^{bb} obtained from the (1+1)-POI fit [170]. 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.

2514 7.5 Conclusion

2515 The analysis of the associated production of vector bosons with boosted Higgs bosons
2516 decaying to a pair of b -quarks using large- R jets is presented. The Higgs candidate is
2517 reconstructed as a large- R jet in order to improve sensitivity in the boosted regime
2518 in which the Higgs decay products are significantly collimated. The analysis is
2519 performed using 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13\text{ TeV}$ collected
2520 throughout the duration of Run 2 of the LHC.

2521 In comparison with the null hypothesis, the Standard Model (SM) VH , $H \rightarrow b\bar{b}$
2522 process is found to have an observed significance of 2.1 standard deviations, whereas
2523 the corresponding expected significance is 2.7 standard deviations. The VH , $H \rightarrow b\bar{b}$
2524 process is measured simultaneously with the diboson VZ , $Z \rightarrow b\bar{b}$ process, which
2525 provide a cross-check for the main analysis. The observed (expected) significance for
2526 the diboson process is 5.4 (5.7).

2527 The statistical and systematic sources of uncertainty contribute a similar amount
2528 to the overall uncertainty on the result. This analysis would therefore likely benefit
2529 greatly from the improved b -tagging efficiency at high- p_{T} enabled by GN1 as discussed
2530 in Chapter 6, due to the associated reduction in statistical uncertainty provided by
2531 the increased number of events used in the analysis.

2532 Chapter 8

2533 Conclusion

2534 Recap why flavour tagging at high pT is important.

2535 Recap challenges

2536 • tracking suffers from several problems which might be improved by...

2537 • tracking might be improved by...

2538 algorithmic gains in flavour tagging can improve things. future work on that front:

2539 • use more info for flavour tagging (hits, calo info, leptons)

2540 • more aux tasks, energy, mass, lxy regression

2541 ultimately analyses will benefit greatly, in particualr boosted xbb tagging and dihiggs

2542 analysis

₂₅₄₃

Bibliography

- ₂₅₄₄ [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- ₂₅₄₅
- ₂₅₄₆ [2] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-027 (2022).
- ₂₅₄₇
- ₂₅₄₈ [3] L. Morel, Z. Yao, P. Cladé and S. Guellati-Khélifa, *Determination of the fine-structure constant with an accuracy of 81 parts per trillion*, *Nature* **588** (2020) pp. 61–65.
- ₂₅₄₉
- ₂₅₅₀
- ₂₅₅₁ [4] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al., *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* **606** (2022) pp. 479–483.
- ₂₅₅₂
- ₂₅₅₃
- ₂₅₅₄ [5] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- ₂₅₅₅
- ₂₅₅₆ [6] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) pp. 2632–2637 [[hep-ex/9503003](#)].
- ₂₅₅₇
- ₂₅₅₈ [7] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) pp. 252–255.
- ₂₅₅₉
- ₂₅₆₀ [8] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* **122** (1983) pp. 103–116.
- ₂₅₆₁
- ₂₅₆₂
- ₂₅₆₃ [9] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) pp. 218–224 [[hep-ex/0012035](#)].
- ₂₅₆₄
- ₂₅₆₅ [10] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- ₂₅₆₆
- ₂₅₆₇ [11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.*

- 2568 **13** (1964) pp. 508–509.
- 2569 [12] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*
2570 *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 2571 [13] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
2572 *Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)
2573 p. 1 [[1207.7214](#)].
- 2574 [14] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*
2575 *experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 2576 [15] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)
2577 p. 083C01.
- 2578 [16] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge*
2579 *invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 2580 [17] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)
2581 pp. 579–588.
- 2582 [18] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 2583 [19] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel*
2584 *symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**
2585 (1968) pp. 367–377.
- 2586 [20] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,
2587 *Phys. Rev.* **104** (1956) pp. 254–258.
- 2588 [21] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental*
2589 *test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 2590 [22] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of*
2591 *conservation of parity and charge conjugation in meson decays: the magnetic moment*
2592 *of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 2593 [23] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,
2594 *Phys. Rev.* **117** (1960) pp. 648–663.
- 2595 [24] J. Goldstone, *Field theories with «superconductor» solutions*, *Il Nuovo Cimento*
2596 (1955–1965) **19** (1961) pp. 154–164.
- 2597 [25] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs*

- 2598 *Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](#).
- 2599 [26] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2600 *ATLAS detector*, [ATLAS-CONF-2018-036](#) (2018).
- 2601 [27] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys.*
2602 *Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](#)].
- 2603 [28] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 2604 [29] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of*
2605 *Instrumentation* **3** (2008) pp. S08002–S08002.
- 2606 [30] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008)
2607 p. S08004.
- 2608 [31] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of*
2609 *Instrumentation* **3** (2008) pp. S08005–S08005.
- 2610 [32] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
2611 *JINST* **3** (2008) p. S08003.
- 2612 [33] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data*
2613 *taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 2615 [34] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*,
2616 (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 2618 [35] CERN, “The accelerator complex.” (2012).
- 2619 [36] G. C. Strong, *On the impact of selected modern deep-learning techniques to the*
2620 *performance and celerity of classification models in an experimental high-energy*
2621 *physics use case*, [2002.01427](#).
- 2622 [37] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022).
2623 <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoveryview/>.
- 2624 [38] ATLAS Collaboration, *Performance of b -jet identification in the ATLAS experiment*,
2625 *JINST* **11** (2016) p. P04008 [[1512.01094](#)].
- 2626 [39] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV*
2627 *using the ATLAS detector at the LHC*, [ATLAS-CONF-2019-021](#) (2019).

- 2628 [40] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design*
2629 *Report, Volume 1*, .
- 2630 [41] for the ATLAS Collaboration collaboration, *Status of the ATLAS Experiment*, Tech.
2631 Rep., Geneva (2010). 10.3204/DESY-PROC-2010-04/1.
- 2632 [42] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design*
2633 *Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 2634 [43] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design*
2635 *Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 2636 [44] ATLAS Collaboration, *The inner detector*, (2022).
2637 <https://atlas.cern/Discover/Detector/Inner-Detector>.
- 2638 [45] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*,
2639 ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 2640 [46] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,
2641 *JINST* **13** (2018) p. T05008 [[1803.00844](https://arxiv.org/abs/1803.00844)].
- 2642 [47] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor*
2643 *tracker*, *JINST* **9** (2014) p. P08009 [[1404.7473](https://arxiv.org/abs/1404.7473)].
- 2644 [48] ATLAS Collaboration, *Calorimeter*, (2022).
2645 <https://atlas.cern/Discover/Detector/Calorimeter>.
- 2646 [49] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, .
- 2647 [50] G. Artoni, *Muon momentum scale and resolution in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ in*
2648 *ATLAS*, Tech. Rep., Geneva (2016). 10.1016/j.nuclphysbps.2015.09.453.
- 2649 [51] ATLAS Collaboration, *Muon spectrometer*, (2022).
2650 <https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- 2651 [52] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur.*
2652 *Phys. J. C* **77** (2017) p. 317 [[1611.09661](https://arxiv.org/abs/1611.09661)].
- 2653 [53] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*,
2654 *ATL-SOFT-PUB-2021-001* (2021).
- 2655 [54] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts,*
2656 *Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 2657 [55] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense*

- 2658 *Environments*, ATL-PHYS-PUB-2015-006 (2015).
- 2659 [56] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in*
2660 *dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) p. 673 [[1704.07983](#)].
- 2661 [57] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon*
2662 *pixel detector*, *JINST* **9** (2014) p. P09009 [[1406.7690](#)].
- 2663 [58] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS*
2664 *Experiment*, .
- 2665 [59] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment*
2666 *in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332
2667 [[1611.10235](#)].
- 2668 [60] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for*
2669 *LHC Run 3*, ATL-PHYS-PUB-2019-015 (2019).
- 2670 [61] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency*
2671 *measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79**
2672 (2019) p. 970 [[1907.05120](#)].
- 2673 [62] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP*
2674 **04** (2008) p. 063 [[0802.1189](#)].
- 2675 [63] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur.Phys.J. C***72** (2012)
2676 p. 1896 [[1111.6097](#)].
- 2677 [64] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*
2678 *performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- 2679 [65] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*
2680 *the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 2681 [66] ATLAS Collaboration, *Jet energy scale measurements and their systematic*
2682 *uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
2683 *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].
- 2684 [67] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a*
2685 *New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008)
2686 p. 242001 [[0802.2470](#)].
- 2687 [68] ATLAS collaboration, *Measurement of the associated production of a Higgs boson*
2688 *decaying into b-quarks with a vector boson at high transverse momentum in pp*

- 2689 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,* *Phys. Lett. B* **816** (2021)
2690 p. 136204 [[2008.02508](#)].
- 2691 [69] ATLAS collaboration, *Performance of jet substructure techniques for large- R jets in*
2692 *proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector,* *JHEP* **09** (2013)
2693 p. 076 [[1306.4945](#)].
- 2694 [70] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in*
2695 *13 TeV proton–proton collisions with the ATLAS detector,* *Eur. Phys. J. C* **79** (2019)
2696 p. 135 [[1807.09477](#)].
- 2697 [71] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R,* *JHEP* **06** (2009) p. 059
2698 [[0903.0392](#)].
- 2699 [72] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
2700 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS,* **ATL-PHYS-PUB-2017-010**
2701 (2017).
- 2702 [73] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the*
2703 *Gaussian Sum Filter-based model for bremsstrahlung,* **ATLAS-CONF-2012-047** (2012).
- 2704 [74] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*
2705 *using the 2015 LHC proton–proton collision data,* **ATLAS-CONF-2016-024** (2016).
- 2706 [75] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in*
2707 *proton–proton collision data at $\sqrt{s} = 13$ TeV,* *Eur. Phys. J. C* **76** (2016) p. 292
2708 [[1603.05598](#)].
- 2709 [76] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
2710 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV,* *Eur. Phys.*
2711 *J. C* **78** (2018) p. 903 [[1802.08168](#)].
- 2712 [77] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*
2713 *b-tagging algorithms for the 2017-18 LHC run,* **ATL-PHYS-PUB-2017-013** (2017).
- 2714 [78] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*
2715 *ATLAS detector,* **ATL-PHYS-PUB-2017-011** (2017).
- 2716 [79] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
2717 *Neural Networks at the ATLAS Experiment,* **ATL-PHYS-PUB-2017-003** (2017).
- 2718 [80] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter*
2719 *Flavour Tagging in ATLAS,* **ATL-PHYS-PUB-2020-014** (2020).

- 2720 [81] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000)
2721 pp. 577–606 [[hep-ph/9912292](#)].
- 2722 [82] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**
2723 (2018) p. 030001.
- 2724 [83] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom*
2725 *and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets*,
2726 [ATL-PHYS-PUB-2014-008](#) (2014).
- 2727 [84] Nazar Bartosik, *Diagram showing the common principle of identification of jets*
2728 *initiated by b-hadron decays*, (2022).
2729 https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png.
- 2730 [85] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).
2731 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2022-03/>.
- 2732 [86] S. Adorni Braccesi Chiassi, *Vector Boson Tagging in the Context of the Search for*
2733 *New Physics in the Fully Hadronic Final State*, Ph.D. thesis, Geneva U., 2021.
2734 [10.13097/archive-ouverte/unige:158558](https://archive-ouverte.unige:158558).
- 2735 [87] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*
2736 *Process Interface for Event Generators*, *ArXiv High Energy Physics - Phenomenology*
2737 *e-prints* (2001) .
- 2738 [88] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*
2739 *standard format for Les Houches Event Files*, *Computer Physics Communications*
2740 **176** (2007) pp. 300–304.
- 2741 [89] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*
2742 *universal approximators*, *Neural Networks* **2** (1989) pp. 359–366.
- 2743 [90] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*
2744 *back-propagating errors*, *nature* **323** (1986) pp. 533–536.
- 2745 [91] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*
2746 *activity*, *The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.
- 2747 [92] J. Hopfield, *Neural networks and physical systems with emergent collective*
2748 *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*
2749 *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 2750 [93] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*

- 2751 networks: *Tricks of the trade*, pp. 9–48, Springer, (2012).
- 2752 [94] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv
2753 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2754 [95] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,
2755 *Journal of High Energy Physics* **2004** (2004) p. 040–040.
- 2756 [96] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*
2757 *carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* **2007** (2007)
2758 p. 126–126.
- 2759 [97] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*
2760 *shower simulations: the powheg method*, *Journal of High Energy Physics* **2007** (2007)
2761 p. 070–070.
- 2762 [98] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*
2763 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*
2764 *Physics* **2010** (2010) .
- 2765 [99] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015)
2766 p. 040 [[1410.8849](#)].
- 2767 [100] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
2768 *ATL-PHYS-PUB-2016-020* (2016).
- 2769 [101] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An*
2770 *introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159
2771 [[1410.3012](#)].
- 2772 [102] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,
2773 *ATL-PHYS-PUB-2014-021* (2014).
- 2774 [103] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013)
2775 p. 244 [[1207.1303](#)].
- 2776 [104] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*
2777 **462** (2001) p. 152.
- 2778 [105] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**
2779 (2010) p. 823 [[1005.4568](#)].
- 2780 [106] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.*
2781 *Instrum. Meth. A* **506** (2003) p. 250.

- 2782 [107] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS*
2783 *detector*, [ATLAS-CONF-2014-018](#) (2014).
- 2784 [108] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2785 *ATLAS detector*, [Phys. Lett. B 786](#) (2018) p. 59 [[1808.08238](#)].
- 2786 [109] ATLAS Collaboration, *Observation of Higgs boson production in association with a*
2787 *top quark pair at the LHC with the ATLAS detector*, [Phys. Lett. B 784](#) (2018) p. 173
2788 [[1806.00425](#)].
- 2789 [110] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*
2790 *using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, [JHEP 03](#)
2791 (2020) p. 145 [[1910.08447](#)].
- 2792 [111] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp*
2793 *collision dataset*, [2211.16345](#).
- 2794 [112] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,
2795 *Secondary vertex finding in jets with neural networks*, [The European Physical Journal](#)
2796 [C 81](#) (2021) .
- 2797 [113] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,
2798 *Set2Graph: Learning Graphs From Sets*, [arXiv e-prints](#) (2020) p. arXiv:2002.08772
2799 [[2002.08772](#)].
- 2800 [114] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,
2801 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,
2802 [arXiv e-prints](#) (2018) p. arXiv:1806.01261 [[1806.01261](#)].
- 2803 [115] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,
2804 *Deep Sets*, [arXiv e-prints](#) (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 2805 [116] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13\text{ TeV}$ data*,
2806 [ATL-PHYS-PUB-2015-037](#) (2015).
- 2807 [117] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*
2808 *experiment using the 2015 and 2016 LHC proton–proton collision data at*
2809 *$\sqrt{s} = 13\text{ TeV}$* , [Eur. Phys. J. C 79](#) (2019) p. 639 [[1902.04655](#)].
- 2810 [118] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*
2811 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, [arXiv e-prints](#) (2020)
2812 p. arXiv:2007.08294 [[2007.08294](#)].

- 2813 [119] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,
2814 *arXiv e-prints* (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 2815 [120] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al.,
2816 *Attention Is All You Need*, *arXiv e-prints* (2017) p. arXiv:1706.03762 [[1706.03762](#)].
- 2817 [121] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv
e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2819 [122] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).
2820 <https://github.com/onnx/onnx>.
- 2821 [123] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs
Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].
- 2823 [124] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson
decaying into b-quarks with a vector boson at high transverse momentum in pp
collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
2824 p. 136204 [[2008.02508](#)].
- 2826
- 2827 [125] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated
Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039
2828 [[1512.02572](#)].
- 2830 [126] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in
proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
2831 p. 076 [[1306.4945](#)].
- 2833 [127] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO
calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010)
2834 p. 043 [[1002.2581](#)].
- 2836 [128] ATLAS collaboration, *Measurement of the Z/γ^* boson transverse momentum
distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **09**
2837 (2014) p. 145 [[1406.3660](#)].
- 2839 [129] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al.,
2840 *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* **72** (2012) p. 1889
2841 [[1111.2034](#)].
- 2842 [130] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*,
2843 *JHEP* **10** (2012) p. 155 [[1206.3572](#)].

- 2844 [131] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, *JHEP* **10** (2013) p. 083 [[1306.2542](#)].
- 2845
- 2846
- 2847 [132] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, *Phys. Rev. D* **68** (2003) p. 073003 [[hep-ph/0306234](#)].
- 2848
- 2849
- 2850 [133] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) pp. 149–156 [[hep-ph/0307206](#)].
- 2851
- 2852
- 2853 [134] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.* **107** (2011) p. 152003 [[1107.1164](#)].
- 2854
- 2855
- 2856 [135] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) p. 1868 [[1111.0761](#)].
- 2857
- 2858 [136] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated WH production and decay at the LHC*, *JHEP* **04** (2014) p. 039 [[1312.1669](#)].
- 2859
- 2860 [137] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron colliders: The fully differential NNLO QCD calculation*, *Phys. Lett. B* **740** (2015) pp. 51–55 [[1407.4747](#)].
- 2861
- 2862
- 2863 [138] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson at NNLO*, *JHEP* **06** (2016) p. 179 [[1601.00658](#)].
- 2864
- 2865 [139] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke, *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) p. 078 [[1211.5015](#)].
- 2866
- 2867
- 2868 [140] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) p. 065 [[1503.01656](#)].
- 2869
- 2870 [141] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) p. 082 [[1410.0217](#)].
- 2871
- 2872 [142] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) p. 023 [[1307.8122](#)].
- 2873
- 2874 [143] O. Brein, R. V. Harlander and T. J. E. Zirke, *$vh@nnlo$ – Higgs Strahlung at hadron*

- 2875 *colliders*, *Comput. Phys. Commun.* **184** (2013) pp. 998–1003 [[1210.5347](#)].
- 2876 [144] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) p. 126 [[0707.3088](#)].
- 2877
2878 [145] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) p. 2930 [[1112.5675](#)].
- 2880
2881 [146] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) p. 111 [[0907.4076](#)].
- 2882
2883
2884 [147] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, *Phys. Rev. D* **81** (2010) p. 054028 [[1001.5034](#)].
- 2885
2886 [148] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, *Phys. Rev. D* **83** (2011) p. 091503 [[1103.2792](#)].
- 2887
2888 [149] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) p. 1547 [[1009.2450](#)].
- 2889
2890 [150] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-* , *Phys. Rev. D* **82** (2010) p. 054018 [[1005.4451](#)].
- 2891
2892 [151] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al., *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) p. 007 [[0811.4622](#)].
- 2893
2894 [152] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) p. 034 [[1905.09127](#)].
- 2895
2896 [153] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) p. 111601 [[1111.5206](#)].
- 2897
2898 [154] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) p. 039 [[0808.3674](#)].
- 2899
2900 [155] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, *JHEP* **03** (2008) p. 038 [[0709.1027](#)].
- 2901
2902 [156] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, *JHEP* **04** (2013) p. 027 [[1207.5030](#)].
- 2903
2904 [157] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson*

- 2905 *Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys.*
2906 *Rev. Lett.* **103** (2009) p. 082001 [[0903.2120](#)].
- 2907 [158] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43**
2908 (2016) p. 023001 [[1510.03865](#)].
- 2909 [159] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to*
2910 *Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* **03**
2911 (2012) p. 075 [[1112.5142](#)].
- 2912 [160] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo*
2913 *program for Higgs production in vector-boson fusion and Higgs strahlung at hadron*
2914 *colliders*, *Comput. Phys. Commun.* **195** (2015) pp. 161–171 [[1412.5390](#)].
- 2915 [161] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in*
2916 *Run-2*, *ATL-PHYS-PUB-2015-022* (2015).
- 2917 [162] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS*
2918 *detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 2919 [163] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using*
2920 *ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, *ATLAS-CONF-2018-006*
2921 (2018).
- 2922 [164] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using*
2923 *a likelihood approach with the ATLAS detector*, *ATLAS-CONF-2018-001* (2018).
- 2924 [165] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the*
2925 *ATLAS Detector*, *ATL-PHYS-PUB-2014-013* (2014).
- 2926 [166] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark*
2927 *pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) p. 836
2928 [[1906.11005](#)].
- 2929 [167] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton*
2930 *distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) p. 204
2931 [[1412.3989](#)].
- 2932 [168] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton*
2933 *distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.*
2934 *D93* (2016) p. 033006 [[1506.07443](#)].
- 2935 [169] J. K. Anders and M. D’Onofrio, *V+Jets theoretical uncertainties estimation via a*

- 2936 *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 2937 [170] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in*
2938 *association with a vector boson and decaying to a pair of b-quarks using large-R jets*,
2939 ATL-COM-PHYS-2019-1125 (2019).
- 2940 [171] J. Pumplin et al., *New generation of parton distributions with uncertainties from*
2941 *global qcd analysis*, *JHEP* **07** (2002) p. 012 [[0201195](#)].
- 2942 [172] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The*
2943 *roostats project*, *arXiv preprint arXiv:1009.1003* (2010) .

2944 List of figures

<small>2945</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. The SM Higgs mechanism relies on a complex scalar doublet with a corresponding 5-dimensional potential.	19
<small>2950</small>	2.2 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.	25
<small>2952</small>	2.3 Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [25]. 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.	25
<small>2957</small>	2.4 Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [25]. 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.	26
<small>2961</small>	3.1 An overview of the CERN accelerator complex [35]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.	29
<small>2965</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. [36].	30

2969	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. [37].	32
2975	3.4 Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [34]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.	34
2979	3.5 Average pile-up profiles measured by ATLAS during Run 2 [34]. Higher levels of pile-up are planned for Run 3.	34
2981	3.6 A 3D model of the entire ATLAS detector [41]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.	35
2985	3.7 A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks [44].	36
2987	3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [37].	37
2989	3.9 A schematic cross-sectional view of the ATLAS IBL [45].	38
2990	3.10 The ATLAS calorimeters [48]. 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 endcaps.	40
2993	3.11 The ATLAS muon spectrometer [51].	42
2994	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 [56].	46
2998	3.13 A sketch of electron reconstruction using the ATLAS detector [74]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.	50

3001	4.1 The truth b -hadron decay radius L_{xy} as a function of truth transverse 3002 momentum p_T for reconstructed b -jets in Z' events. Error bars show the 3003 standard deviation. The pixel layers are shown in dashed horizontal lines.	54
3004	4.2 Diagram of a typical b -jet (blue) which has been produced in an event 3005 alongside two light jets (grey) [84]. The b -hadron has travelled a significant 3006 distance (pink dashed line) from the primary interaction point (pink dot) 3007 before its decay. The large transverse impact parameter d_0 is a characteristic 3008 property of the trajectories of b -hadron decay products.	55
3009	4.3 At lower p_T (left) the decay length of the b -hadron is reduced, and the 3010 resulting decay tracks are less collimated. At higher p_T (right) the b -hadron 3011 decay length increases and the resulting decay tracks are more collimated and 3012 have less distance over which to diverge before reaching detector elements. 3013 As a result, the ID may be unable to resolve charged depositions from 3014 different particles, resulting merged clusters.	56
3015	4.4 b -hadron decay track reconstruction efficiency as a function of truth b -hadron 3016 p_T [85]. Nominal track reconstruction is shown in black, while the track 3017 reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver 3018 efficiency) is shown in green. For high- p_T b -hadrons, the ambiguity solver 3019 is overly aggressive in its removal of b -hadron decay tracks. Suggestions for 3020 the improvement of the track reconstruction efficiency in this regime by the 3021 loosening of cuts in the ambiguity solver are shown in blue and red. . . .	57
3022	4.5 Hit multiplicities on the IBL (left) and the pixel layers (right) as a function 3023 of the p_T of the reconstructed track. Tracks from the weak decay of 3024 the b -hadron are shown in red, while fragmentation tracks (which are 3025 prompt) are in blue. Baseline tracks are those produced in the standard 3026 reconstruction described in Section 3.4.1, while pseudotracks represent the 3027 ideal performance of the ATLAS detector and are described in Section 3.4.1. 3028 Hit multiplicities on the pseudotracks decrease at high p_T due to the flight 3029 of the b -hadron before its decay. The baseline tracks have more hits than the 3030 pseudotracks, indicating that they are being incorrectly assigned additional 3031 hits on the inner layers of the detector.	58
3032	4.6 The fraction of hits which are shared on b -hadron decay tracks on the IBL 3033 (left), and the B-layer (right), as a function of the production radius of 3034 the b -hadron decay product. Pseudotracks represent the ideal performance 3035 given the ATLAS detector, see Section 3.4.1.	60

3036	4.7 The rate to assign good (left) and wrong (right) IBL hits as a function 3037 of b -hadron p_T for tracks using baseline tracking (black) and the modified 3038 version of the outlier removal procedure (red). For each track, the corre- 3039 sponding p_T bin is filled with the number of good or wrong hits and this 3040 value is averaged to show the overall rate.	62
3041	4.8 (left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified 3042 GX2F tracks. (right) The absolute value of the d_0 pull as a function of 3043 b -hadron transverse momentum.	62
3044	5.1 A diagram displaying the logical flow of a single neuron with three inputs 3045 x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are 3046 summed. A bias term θ_0 is added, and the result z is passed to an activation 3047 function. Each neuron can be thought of as a logistic regression model. .	66
3048	5.2 The output of several common choices for the activation function $g(z)$ of 3049 an artificial neuron. The input z is the output of the dot product between 3050 the activation and the weights, plus a bias term.	67
3051	5.3 Rate of fake tracks as a function of jet transverse momentum (left) and 3052 $\Delta R(\text{track}, \text{jet})$ (right). The rate of fake tracks increases significantly as a 3053 function of p_T , and also increases as the distance to the jet axis decreases.	70
3054	5.4 The light-jet efficiency of the low level tagger SV1 for jets in the Z' sample 3055 with $250 < p_T < 5000 \text{ GeV}$, as a function of b -jet efficiency. The nominal 3056 tracking setup (black) is shown alongside the case where fake tracks which 3057 are not from the decay of a b -hadron are removed. The light-jet efficiency 3058 is decreased, demonstrating that the presence of fake tracks is detrimental 3059 to the algorithm performance.	70
3060	5.5 (left) Normalised histogram of the model output separated for signal and 3061 fake tracks, and further separated by those tracks which are from the 3062 decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and 3063 tracks from the decay of a b -hadron (dashed line). The model is able to 3064 successfully discriminate between signal and fake tracks, and shows a very 3065 similar performance when looking specifically at tracks from the decay of a 3066 b -hadron.	75

3067	5.6 (left) Normalised histogram of the model output separated for tracks from 3068 the decay of a b -hadron and tracks from other sources. The two groups are 3069 further separated into those tracks which are fake. (right) The ROC curve 3070 for all tracks (solid line).	77
3071	5.7 The effect of applying the fake track identification algorithm together with 3072 the b -hadron decay track identification on the jet tagging performance of 3073 SV1 for jets in the Z' sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and 3074 $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) 3075 is compared to two working points of fake track removal, labelled “A” (red) 3076 and “B” (orange), which represent two different 2D working points of the 3077 track classification tools. Removal of fake tracks based on truth information 3078 is shown by the green curves.	78
3079	6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right) [2]. 3080 The existing approach utilises low-level algorithms (shown in blue), the 3081 outputs of which are fed into a high-level algorithm (DL1r). Instead 3082 of being used to guide the design of the manually optimised algorithms, 3083 additional truth information from the simulation is now being used as 3084 auxiliary training targets for GN1. The solid lines represent reconstructed 3085 information, whereas the dashed lines represent truth information.	81
3086	6.2 The flow of information through a graph neural network layer for (left) a full 3087 implementation of the layer and (right) a deep sets model [115]. Reproduced 3088 from Ref. [114].	84
3089	6.3 The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of n_{tracks} , 3090 where each track is described by 21 track features ($n_{\text{tf}} = 21$) [2]. The jet 3091 features are copied for each of the tracks, and the combined jet-track vectors 3092 of length 23 form the inputs of GN1.	89
3093	6.4 The network architecture of GN1. Inputs are fed into a per-track initialisa- 3094 tion network, which outputs an initial latent representation of each track. 3095 These representations are then used to populate the node features of a 3096 fully connected graph network. After the graph network, the resulting node 3097 representations are used to predict the jet flavour, the track origins, and 3098 the track-pair vertex compatibility.	89

3099	6.5 Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample [2]. 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.	95
3100		
3101		
3102		
3103		
3104		
3105	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 [2]. 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.	97
3106		
3107		
3108		
3109		
3110		
3111		
3112	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 [2]. 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.	97
3113		
3114		
3115		
3116		
3117		
3118		
3119	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 [2]. 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 accross the p_T range shown. Binomial error bands are denoted by the shaded regions.	98
3120		
3121		
3122		
3123		
3124		
3125	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 [2]. 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.	100
3126		
3127		
3128		
3129		
3130		

3131	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 [2]. 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.	100
3137	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 [2]. 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.	102
3144	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 [2]. 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.	102
3151	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 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 pileup.	104

3160	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 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.	104
3168	6.15 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.	107
3175	6.16 Inclusive vertex finding efficiency for multitrack truth vertices in b -jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T . 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.	108
3181	6.17 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.	110
3188	6.18 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 DL1d algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions.	112

3194	6.19 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 3195 tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser 3196 track selection trainings of GN1. The ratio to the performance of the DL1d 3197 algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in 3198 the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial 3199 error bands are denoted by the shaded regions.	112
3200	7.1 Diagrams of the signal process (top) and the 0-lepton and 1-lepton $t\bar{t}$ back- 3201 grounds (middle, bottom). Object to the right of centre are reconstructed 3202 within the large- R jet. For the backgrounds, the large- R jet contains a 3203 mis-tagged c - or light-jet.	122
3204	7.2 Normalised leading large- R jet mass distribution from Z and $W+hf$ pro- 3205 cesses in the HP SR of the 0-lepton channel [170]. The renormalisation scale 3206 μ_r has been varied by a factor of 2 (1up) and 0.5 (1down). An exponential 3207 function is fitted to the ratio between the nominal and variation samples.	132
3208	7.3 The comparison on m_J shapes between SHERPA and MADGRAPH samples 3209 from $V+hf$ process in p_T^V inclusive signal regions [170]. The Kolmogorov- 3210 Smirnov test and χ^2/ndf are shown on the plots.	133
3211	7.4 Normalised leading large- R jet mass distribution from WZ and ZZ pro- 3212 cesses, merged among all the signal regions and lepton channels [170]. The 3213 renormalisation scale μ_R has been varied by a factor of 2 (1up) and 0.5 3214 (1down). The red line shape shows the fitting results of the hyperbolic 3215 tangent function.	135
3216	7.5 The comparison on m_J shapes between SHERPA and POWHEG+PYTHIA 8 3217 samples from WZ and ZZ process in high and low purity signal regions [170]. 3218 p_T^V regions and 0- and 1-lepton channels are merged. The dashed green line 3219 shows the fitted third order polynomial function and the blue lines show 3220 the function after a protection is added in the high mass region.	136

3221	7.6 The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 3222 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$ 3223 [124]. The LP and HP regions are merged for the 0-lepton and 1-lepton 3224 channels. The fitted background contributions are shown as filled histograms. 3225 The Higgs boson signal ($m_H = 125 \text{ GeV}$) is shown as a filled histogram and 3226 is normalised to the signal yield extracted from data ($\mu_{VH}^{bb} = 0.72$), and 3227 as an unstacked unfilled histogram, scaled by the SM prediction times a 3228 factor of two. The size of the combined on the sum of the fitted signal and 3229 background is shown in the hatched band. The highest bin contains the 3230 overflow.	142
3231	7.7 The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton 3232 channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) 3233 the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$ [124]. 3234 The background contributions after the likelihood fit are shown as filled 3235 histograms. The Higgs boson signal ($m_h = 125 \text{ GeV}$) is shown as a filled 3236 histogram on top of the fitted backgrounds normalised to the signal yield 3237 extracted from data ($\mu_{VH}^{bb} = 0.72$), and unstacked as an unfilled histogram, 3238 scaled by the SM prediction times a factor of 2. The size of the combined 3239 statistical and systematic uncertainty for the sum of the fitted signal and 3240 background is indicated by the hatched band. The highest bin in the 3241 distributions contains the overflow.	143
3242	7.8 m_J distribution in data after subtraction of all backgrounds except for the 3243 WZ and ZZ diboson processes [124]. The contributions from all lepton 3244 channels and signal regions are summed and weighted by their respective 3245 values of the ratio of fitted Higgs boson signal and background yields. The 3246 expected contribution of the associated WH and ZH production of a SM 3247 Higgs boson with $m_H = 125 \text{ GeV}$ is shown scaled by the measured combined 3248 signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to 3249 its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and 3250 systematic uncertainty is indicated by the hatched band. This error band is 3251 computed from a full signal-plus-background fit including all the systematic 3252 uncertainties defined in Section 7.2, except for the VH/VZ experimental 3253 and theory uncertainties.	144

3254	7.9 Signal strength compatibility test between the (3+1)-POI fit (with the 3255 three lepton channels splitted) and the default (1+1)-POI fit [124]. The 3256 compatibility of the three channels is evaluated via a χ^2 difference test and 3257 results in a p-value of 49%.	145
3258	7.10 Impact of systematic uncertainties on the fitted VH signal-strength parameter 3259 μ_{VH}^{bb} sorted in decreasing order [170]. The boxes show the variations 3260 of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual 3261 nuisance parameter to its post-fit value modified upwards or downwards by 3262 its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of 3263 up- and down-variations can be distinguished via the dashed and plane box 3264 fillings. The yellow boxes show the pre-fit impact (top x -axis) by varying 3265 each nuisance parameter by ± 1 . The filled circles show the deviation of the 3266 fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value 3267 θ_0 expressed in standard deviations with respect to their nominal uncer- 3268 tainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainties 3269 on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show 3270 the fitted values and uncertainties of the normalization parameters that are 3271 freely floating in the fit. Pre-fit, these parameters have a value of one.	147

3272 List of tables

<small>3273</small>	2.1 The fermions of the SM [15]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.	<small>14</small>
<small>3276</small>	2.2 The bosons of the SM [15]. 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.	<small>15</small>
<small>3280</small>	3.1 Overview of the different LHC runs [33, 34]. 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. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.	<small>28</small>
<small>3284</small>	5.1 Truth origins which are used to categorise the physics process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [56]. A truth-matching probability of less than 0.5 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.	<small>69</small>
<small>3291</small>	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 [56], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.	<small>73</small>

3297	5.3 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). Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [56]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.	74
3303	5.4 Hyperparameter for the track classification model	74
3304	5.5 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. The continuous output of the model allows for the tuning of good and fake track identification efficiencies.	76
3308	5.6 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).	77
3312	6.1 Input features to the GN1 model [2]. 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 [56], 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.	86
3320	6.2 A summary of GN1’s different classification networks used for the various training objectives, adapted from Ref. [2]. The hidden layers column contains a list specifying the number of neurons in each layer.	91
3323	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.	109

3330	7.1 Summary of the definitions of the different analysis regions [124]. 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 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.	116
3331		
3332		
3333		
3334		
3335		
3336	7.2 Signal and background processes with the corresponding generators used for the nominal samples [124]. 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 [158] 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 [159, 160].	118
3337		
3338		
3339		
3340		
3341		
3342		
3343		
3344		
3345		
3346		
3347		
3348	7.3 Event selection requirements for the boosted VH , $H \rightarrow b\bar{b}$ analysis channels and sub-channels [124]. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region.	120
3349		
3350		
3351		
3352	7.4 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.	126
3353		
3354		
3355	7.5 0-lepton $W+jets$ nominal sample flavour composition and total event yield [170].	128
3356	7.6 1-lepton $W+jets$ nominal sample flavour composition and total event yield [170].	128
3357	7.7 0-lepton $Z+jets$ nominal sample flavour composition and total event yield [170].	128
3358	7.8 2-lepton $Z+jets$ nominal sample flavour composition and total event yield [170].	128

3359	7.9 V+jets acceptance uncertainties [170]. W +jets SR and CR uncertainties marked with a superscript \dagger are correlated. The 1L W +jets H/M uncertainty marked by * is applied as independent and uncorrelated NPs in both HP and LP signal regions. The 0L W +jets Wbb Norm uncertainty is only applied when a floating normalisation for Wbb cannot be obtained from the 1L channel. A 30% uncertainty for $Z \rightarrow b\bar{b}$ norm is applied in the 1L channel when a floating normalisation for $Z \rightarrow b\bar{b}$ cannot be obtained from the 0L or 2L channels.	130
3367	7.10 Diboson acceptance uncertainties [170]. All uncertainties except channel ex- trapolation uncertainties are fully correlated between ZZ and WZ processes and channels.	135
3370	7.11 Factors applied to the nominal normalisations of the $t\bar{t}$, W +hf, and Z +hf backgrounds, as obtained from the likelihood fit [124]. The errors represent the combined statistical and systematic uncertainties.	140
3373	7.12 Breakdown of the observed absolute contributions to the uncertainty on the signal strength μ_{VH}^{bb} obtained from the (1+1)-POI fit [170]. The average impact represents the average between the positive and negative uncer- tainties 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.	148