

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**

December 15, 2022

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

Here is an example of how to declare commands for use in a single file that will not be needed elsewhere. Additionally, it serves to illustrate the chapter referencing system.

J. Shlomi for GNN.

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

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 b-tagging	52
4.1	b -hadron Reconstruction	53
4.1.1	Decay Topology	53
4.1.2	Challenges Facing b -hadron Reconstruction	55
4.2	Investigations into High p_T b -hadron Tracking	58
4.2.1	Shared Hits	59
4.2.2	Global χ^2 Fitter Outlier Removal	60
4.3	Conclusion	63
5	Track Classification MVA	65
5.1	Machine Learning Background	65
5.1.1	Neural Networks	66
5.1.2	Training with Gradient Descent	68
5.2	Track Truth Origin Labelling	69
5.3	Fake Track Identification Tool	70
5.3.1	Datasets	72
5.3.2	Model Inputs	73
5.3.3	Model Hyperparameters	73
5.3.4	Results	76
5.4	b -hadron Track Identification	77
5.5	Combined Approach	77
5.6	Conclusion	79
6	Graph Neural Network Flavour Tagger	82
6.1	Motivation	83
6.2	Graph Neural Network Theory	86
6.3	Experiemental Setup	87
6.3.1	Datasets	87
6.4	Model Architecture	88
6.4.1	Model Inputs	88
6.4.2	Auxiliary Training Objectives	89
6.4.3	Architecture	91
6.4.4	Training	95

6.5	Results	96
6.5.1	<i>b</i> -tagging Performance	97
6.5.2	<i>c</i> -tagging Performance	102
6.5.3	Jet Display Diagrams	103
6.5.4	Ablations	105
6.5.5	Inclusion of Low-Level Vertexing Algorithms	107
6.5.6	Vertexing Performance	110
6.5.7	Track Classification Performance	113
6.5.8	Looser Track Selection	116
6.6	Conclusion	116
7	Boosted VHbb Analysis	121
7.1	Analysis Overview	122
7.1.1	Data & Simulated Samples	123
7.1.2	Object Reconstruction	124
7.1.3	Selection Criteria	126
7.1.4	Control Regions	128
7.1.5	Background Composition	128
7.2	Systematic Uncertainties & Background Modelling	130
7.2.1	Sources of Systematic Uncertainties	131
7.2.2	Implementation of Variations	133
7.2.3	Vector Boson + Jets Modelling	134
7.2.4	Diboson Modelling	141
7.3	Statistical Treatment	143
7.3.1	Likelihood Function	144
7.3.2	Background Normalisations	146
7.3.3	Asimov Dataset & Expected Results	147
7.4	Results	147
7.4.1	Post-fit Distributions	148
7.4.2	Observed Signal Strength & Significance	148
7.4.3	Impact of Systematics	152
7.5	Improved <i>b</i> -tagging using GNNs	155
7.6	Conclusion	155
8	Conclusion	157

Bibliography 158

List of figures 172

List of tables 183

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 also played a key
36 role in software r22 validation studies for the Flavour Tagging group, including the
37 validation of the quasi-stable particle interaction simulation. The author maintains
38 and contributes to various software frameworks used in the Flavour Tagging group,
39 and contributes to group documentation.
- 40 **Higgs:** The author was an active member of the Boosted VHbb analysis group The
41 author performed various studies deriving systematic uncertainties for the $V+jets$
42 and diboson backgrounds. The author also produced and maintained samples, ran fit
43 studies and cross checks, and gave the diboson unblinding approval talk to the Higgs
44 group. The author also contributed to the developement of the analysis software.

⁴⁵ **Chapter 2**

⁴⁶ **Theoretical Framework**

⁴⁷ The Standard Model (SM) of particle physics is the theory describing all known
⁴⁸ elementary particles and their interactions via three of the four fundamental forces.
⁴⁹ Developed by merging the successful theories of quantum mechanics and relativity
⁵⁰ in the second half of the 20th century, the SM's position today at the centre of our
⁵¹ understanding of the nature of the Universe is firmly established by an unparalleled
⁵² level of agreement between the predictions from the model and experimental results
⁵³ [2, 3].

⁵⁴ The SM has predicted the discovery of the top and bottom quarks [4–6], the W
⁵⁵ and Z bosons [7], and the tau neutrino [8]. The last missing piece of the SM to be
⁵⁶ discovered was the Higgs boson, first theorised in the 1960s [9–11], and eventually
⁵⁷ observed at the LHC in 2012 [12, 13]. After its discovery, much ongoing work has
⁵⁸ been carried out performing detailed measurements of its mass and interactions with
⁵⁹ other particles.

⁶⁰ In this chapter, an overview of the SM is given in Section 2.1, and a more detailed
⁶¹ discussion of the Higgs sector and Higgs phenomenology is provided in Section 2.2.

⁶² **2.1 The Standard Model**

⁶³ The SM is formulated in the language of Quantum Field Theory (QFT). In this
⁶⁴ framework, particles are localised excitations of corresponding quantum fields, which
⁶⁵ are operator-valued distributions across spacetime.

66 Central to QFT is the Lagrangian density which describes the kinematics and
 67 dynamics of a field. Observations of conserved quantities are linked, via Noether's
 68 theorem, to symmetries which are expressed by the Lagrangian. Alongside Global
 69 Poincaré symmetry, the SM Lagrangian observes a local non-Abelian $SU(3)_C \otimes$
 70 $SU(2)_L \otimes U(1)_Y$ gauge symmetry. Gauge symmetries leave observable properties of
 71 the system unchanged when the corresponding gauge transformations are applied
 72 to the fields. The full Lagrangian of the SM can be broken up into distinct terms
 73 corresponding to the different sectors, as in Eq. (2.1). An overview of each sector is
 74 given in the following chapters.

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

75 The SM provides a mathematical description of how three of the four fundamental
 76 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 [14]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.

77

78 2.1.1 Quantum Electrodynamics

79 Quantum electrodynamics (QED) is the relativistic quantum theory which describes
 80 the interaction between the photon and charged matter. Consider a Dirac spinor
 81 field $\psi = \psi(x)$ and its adjoint $\bar{\psi} = \psi^\dagger \gamma^0$, where ψ^\dagger denotes the Hermitian conjugate
 82 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 [14]. 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.

⁸³ Dirac Lagrangian density is

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

⁸⁴ 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

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

⁸⁷ Suppose some fundamental symmetry that requires invariance under a local $U(1)$
⁸⁸ gauge transformation

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

⁸⁹ where α varies over every spacetime point x . Under this transformation, the Dirac
⁹⁰ equation transforms as

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

⁹¹ For the Dirac equation to remain invariant under the transformation in Eq. (2.4),
⁹² a new field A_μ which transforms as $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x)$ must be added. The
⁹³ 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 [15] 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

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

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

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

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

123 The full QCD Lagrangian is then given by

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

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

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

134 2.1.3 The Electroweak Sector

135 The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam
136 (GWS) model of electroweak interaction [16–18]. The Lagrangian is specified by
137 requiring invariance under the symmetry group $SU(2)_L \otimes U(1)_Y$, as motivated by a
138 large amount of experimental data. Here, $SU(2)_L$ is referred to as weak isospin and
139 $U(1)_Y$ as weak hypercharge.

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

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

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

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

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

- 150 The weak force violates parity conservation [19–21], i.e. invariance under parity
151 transformations (mirror reflections). Only left handed fermions participate in the
152 weak interaction. Since there is no other force through which neutrinos interact with
153 other particles, there are no right handed neutrinos in the standard model.

154 2.2 The Higgs Mechanism

- 155 The Brout-Englert-Higgs mechanism (henceforth just “Higgs mechanism”) is the
156 mechanism through which the fundamental particles of the SM acquire mass [9–11].
157 Experimentally it was known that the weak force had a weak effective strength,
158 which was suggestive of a massive mediating gauge particle. However, directly
159 adding mass to the weak gauge bosons violates the non-Abelian symmetry of the
160 SM. Instead, the gauge bosons can gain mass through the interaction with a scalar
161 Higgs field which results from the spontaneous breakdown of symmetry as discussed
162 in Section 2.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as
163 discussed in Section 2.2.2. Section 2.2.3 described some basic phenomenology of the
164 Higgs particle relevant to hadron colliders.

¹⁶⁵ 2.2.1 Electroweak Symmetry Breaking

¹⁶⁶ Spontaneous symmetry breaking (SSB) is a key part of the Higgs mechanism. It
¹⁶⁷ is the transition of a physical system from a state of manifest symmetry to a state
¹⁶⁸ of hidden, or *broken*, symmetry. In particular, this applies to physical systems
¹⁶⁹ where the Lagrangian observes some symmetry, but the lowest energy vacuum states
¹⁷⁰ do not exhibit that same symmetry. In other words, the symmetry is broken for
¹⁷¹ perturbations around the vacuum state.

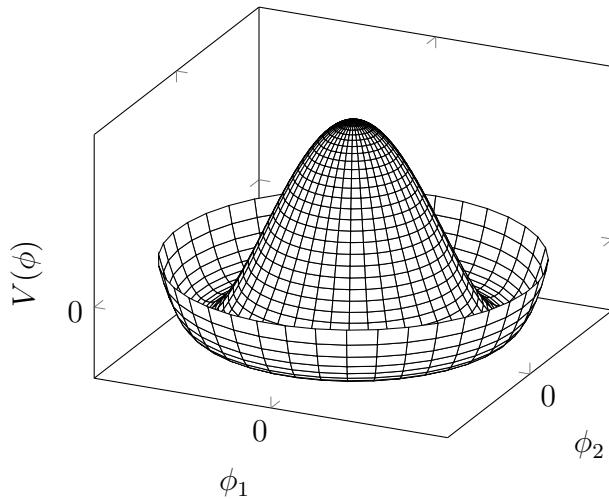


Figure 2.1: The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. The SM Higgs mechanism relies on a complex scalar doublet with a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

¹⁷² Consider gauge fields from the local $SU(2)_L \otimes U(1)_Y$ symmetry group discussed in
¹⁷³ Section 2.1.3 coupled to a complex scalar field $\phi = \phi(x)$. The scalar field ϕ transforms
¹⁷⁴ as a weak isospin doublet. Omitting the kinetic term of the gauge fields, and writing
¹⁷⁵ $\phi^2 \equiv \phi^\dagger \phi$, the Lagrangian is

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

¹⁷⁶ 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 [22, 23]), 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

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

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

204 where θ_W is the weak mixing angle defined by

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

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

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

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

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

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

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

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

229 2.2.2 Fermionic Yukawa Coupling

230 Adding the masses of the fermions by hand breaks the gauge invariance of the
231 theory. Instead, we can use a Yukawa coupling between the fermion fields and the
232 Higgs field in order to generate mass terms after spontaneous electroweak symmetry
233 breakdown [17]. In this way, the fermion masses are determined by both the respective
234 couplings to the Higgs field and the VEV of the Higgs field itself, which sets the
235 basic mass scale of the theory.

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

239 Lepton Masses

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

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

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

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

246 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
247 the first term in Eq. (2.22) which is quadratic in the electron field, and can therefore
248 be identified as the electron mass term. An interaction term between the electron
249 field e and the true vacuum Higgs field H is also present. Mass is generated for the
250 other lepton generations in the same way.

251 **Quark Masses**

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

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

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

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

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

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

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

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

269 where the size of the elements $|V_{pq}|^2$ measures the probability of a transition between
270 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

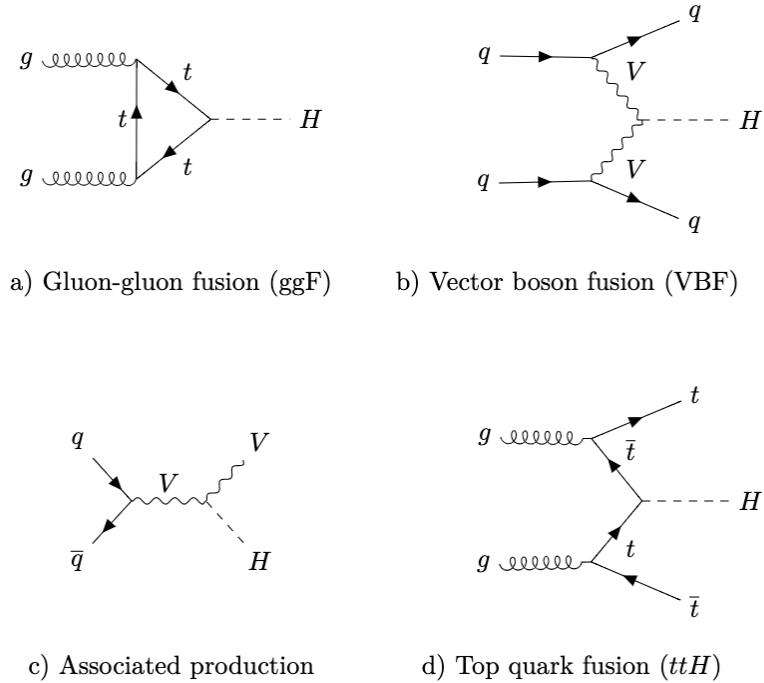


Figure 2.2: Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV.

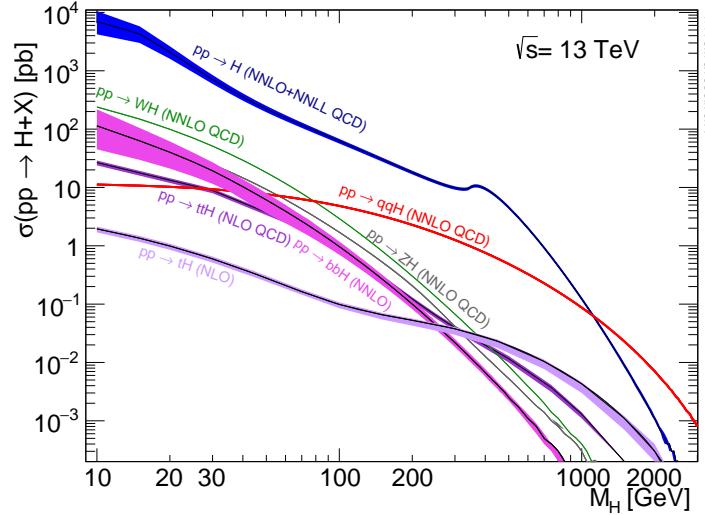


Figure 2.3: Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [24]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, Higgs boson production is dominated by gluon-gluon fussion, vector boson fusion, associated production with vector bosons, and top quark fusion.

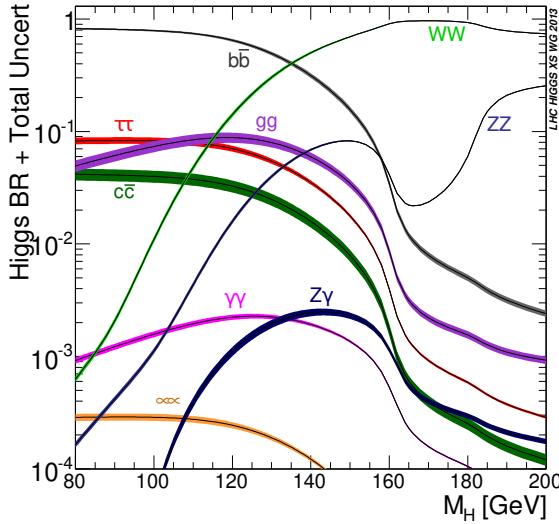


Figure 2.4: Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13 \text{ TeV}$ [24]. Uncertainties are shown in the shaded bands. At $m_H = 125 \text{ GeV}$, the Higgs predominantly decays to a pair of b -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of W bosons.

of b -quarks, i.e. $pp \rightarrow VH(b\bar{b})$. 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 [25, 26]. 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.

309 Chapter 3

310 The Large Hadron Collider and the
311 ATLAS Detector

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

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

³²⁸ 3.1 The Large Hadron Collider

³²⁹ The LHC is operated in multi-year *runs* during which beams of protons are circulated
³³⁰ and collided. Between runs there are periods of shutdown while the accelerator and
³³¹ detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC
³³² collided proton bunches, each containing more than 10^{11} particles, 20 million times
³³³ per second, providing 7 TeV proton-proton collisions at instantaneous luminosities
³³⁴ of up to $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The centre-of-mass energy was increased to 8 TeV
³³⁵ towards the end of Run 1 in 2012. Run 2, which spanned in 2015–2018, further
³³⁶ increased the the proton-proton collision energy to 13 TeV. During Run 2 the bunch
³³⁷ spacing was reduced, leading to a collisison rate of 40 MHz. Over the course of
³³⁸ Run 2 a total usable integrated luminosity of 139 fb^{-1} was recorded. 2022 marked the
³³⁹ beginning of Run 3 which, with a higher center of mass energy and peak luminosity,
³⁴⁰ is expected to culminate in the approximate tripling of the dataset size. A summary
³⁴¹ 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 [32,33]. 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.

³⁴² An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is
³⁴³ at the final stage of a chain of accelerators which incrementally step-up the energy
³⁴⁴ of incoming protons. The first accelerator is Linac4, a linear accelerator which
³⁴⁵ accelerates negative hydrogen ions to an energy of 160 MeV. Upon leaving Linac4,
³⁴⁶ the ions are stripped of both electrons and the resulting protons are fed into the
³⁴⁷ Proton Synchrotron Booster (PSB), which increases the energy of the protons to
³⁴⁸ 2 GeV. The protons leaving the PSB are passed to the Proton Synchrotron (PS),
³⁴⁹ which increases the energy to 26 GeV, and then from the PS to the Super Proton
³⁵⁰ Synchrotron (SPS) which further increases the energy to 450 GeV. Finally, the proton

351 beams are injected in the LHC where they are accelerated to their final energy of
 352 6.5 TeV (for Run 2).

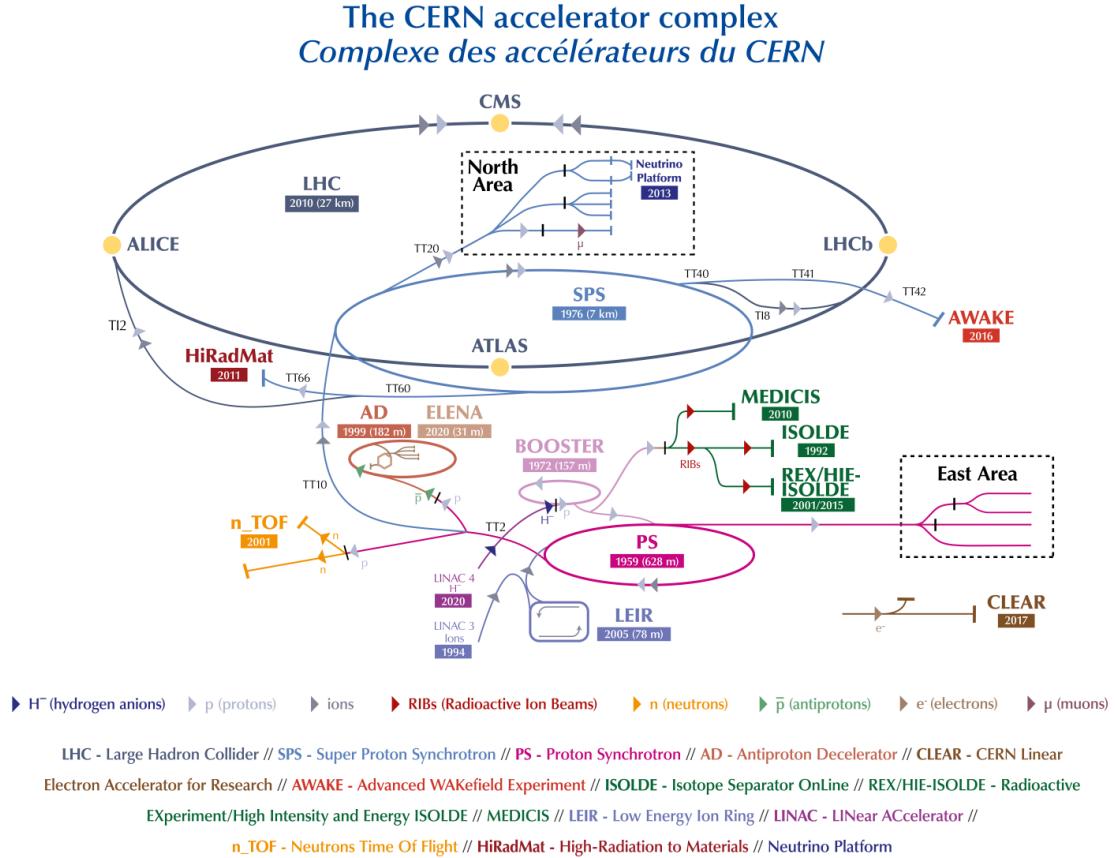


Figure 3.1: An overview of the CERN accelerator complex [34]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

353 3.2 Coordinate System & Collider Definitions

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

358 3.2.1 ATLAS Coordinate System

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

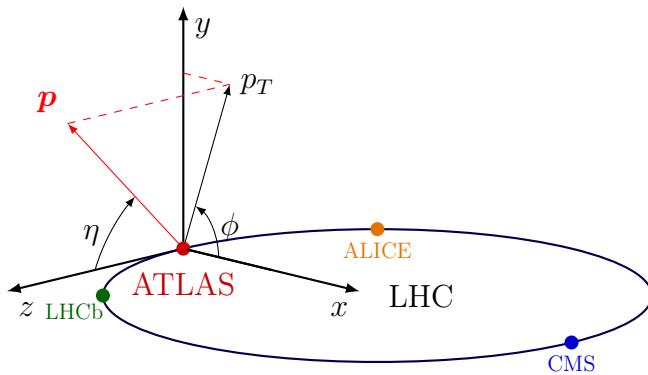


Figure 3.2: The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [35].

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

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

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

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

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

³⁷² Angular distance between two objects is measured in units of ΔR and is defined as
³⁷³ the sum in quadrature of the η and ϕ displacements

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

³⁷⁴ 3.2.2 Track Parameterisation

³⁷⁵ The trajectories of charged particle tracks are parameterised as a helix which is
³⁷⁶ fully specified using five parameters: $(d_0, z_0, \phi, \theta, q/p)$. Transverse and longitudinal
³⁷⁷ impact parameters (IP) d_0 and z_0 specify the closest approach of the trajectory of
³⁷⁸ a particle to the origin. The transverse IP d_0 and longitudinal IP z_0 are measured
³⁷⁹ with respect to the hard scatter primary vertex (see Section 3.4.2). ϕ and θ are
³⁸⁰ the azimuthal and polar angles respectively, and q/p is the measured charge on the
³⁸¹ track¹ divided by the scalar 3-momentum. Fig. 3.3 shows each of these parameters
³⁸² diagrammatically.

³⁸³ Impact parameter significances are defined as the IP divided by its corresponding
³⁸⁴ uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. When used in flavour tagging
³⁸⁵ (see Chapter 4), track IP significances are lifetime signed according to the track's
³⁸⁶ direction with respect to the jet axis and the primary vertex [37]. The sign IP
³⁸⁷ significances is positive if the track crosses the jet axis in front of the primary vertex
³⁸⁸ and negative if the crossing is behind the primary vertex.

³⁸⁹ 3.2.3 Hadron Collider Definitions

³⁹⁰ Cross Section

³⁹¹ The cross section σ is closely related to the probability of an interaction between
³⁹² two colliding particles, and is analogous to an effective cross-sectional area of the
³⁹³ particles. The cross section of a process depends on the transition matrix element,
³⁹⁴ obtained using the Feynman rules of the theory which are derived using QFT, and a
³⁹⁵ 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 .

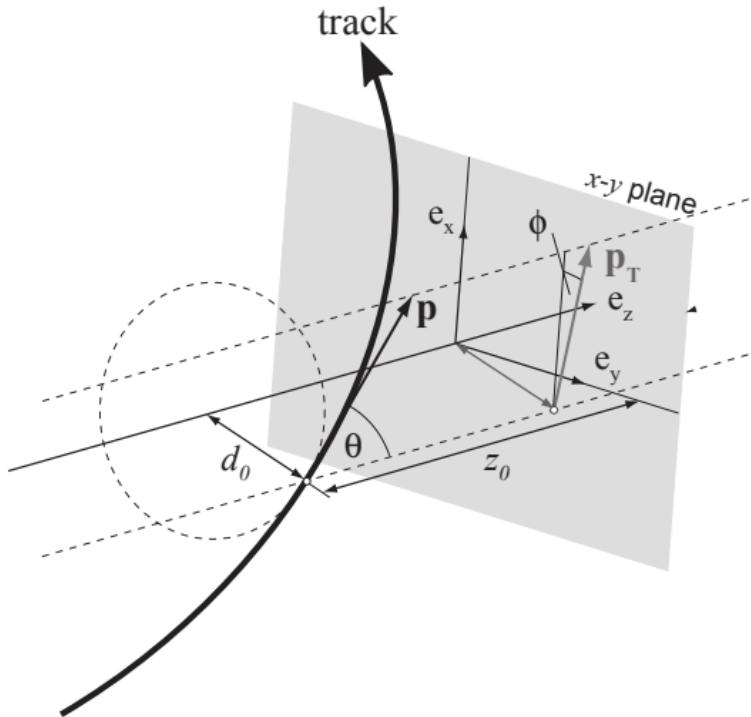


Figure 3.3: The track parameterisation used at the ATLAS detector. Five coordinates $(d_0, z_0, \phi, \theta, q/p)$ are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum \mathbf{p} and 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. [36].

³⁹⁶ section can be factorised as

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

³⁹⁷ The partonic cross section can be calculated at sufficiently high energies such as
³⁹⁸ those found at the LHC, while the parton distribution functions (PDFs) have to be
³⁹⁹ extracted from experimental results.

⁴⁰⁰ Luminosity

⁴⁰¹ The total number of proton-proton collisions N is related to the total pp cross σ
⁴⁰² section by the integrated luminosity L , as in

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

⁴⁰³ The instantaneous luminosity \mathcal{L} relates the cross section to the number of collisions
⁴⁰⁴ per unit time. For two colliding bunched proton beams, it is defined as

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

⁴⁰⁵ where n_1 and n_2 are the number of protons in the colliding bunches, f is the bunch
⁴⁰⁶ crossing frequency, and σ_x and σ_y are the rms width of the beam in the horizontal
⁴⁰⁷ and vertical directions.

⁴⁰⁸ The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In
⁴⁰⁹ total, 139 fb^{-1} of usable physics data was collected over the three-year run. The
⁴¹⁰ uncertainty on the total integrated luminosity is 1.7% [38].

⁴¹¹ Pile-up

⁴¹² At the centre of the ATLAS detector, bunches of more than 10^{11} protons meet at a
⁴¹³ small crossing angle. Each bunch-crossing is called an *event*. There is generally at
⁴¹⁴ most one hard proton-proton scatter per event. Additional interactions are typically
⁴¹⁵ relatively soft and are known as *pile-up*. Pile-up from interactions within the same
⁴¹⁶ bunch-crossing is known as *in-time* pile-up while residual signatures from previous
⁴¹⁷ bunch-crossings is known as *out-of-time* pile-up. The number of pile-up interactions

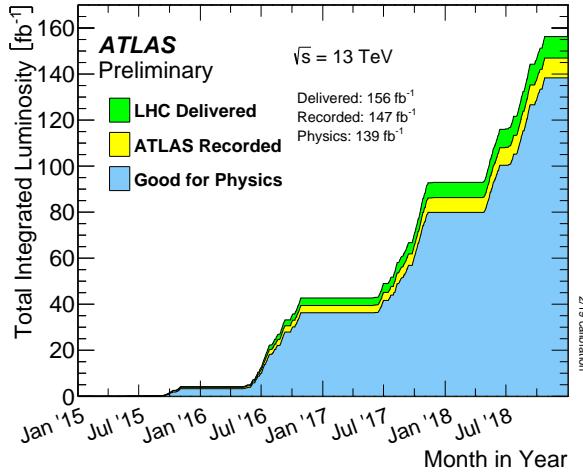


Figure 3.4: Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [33]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.

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.

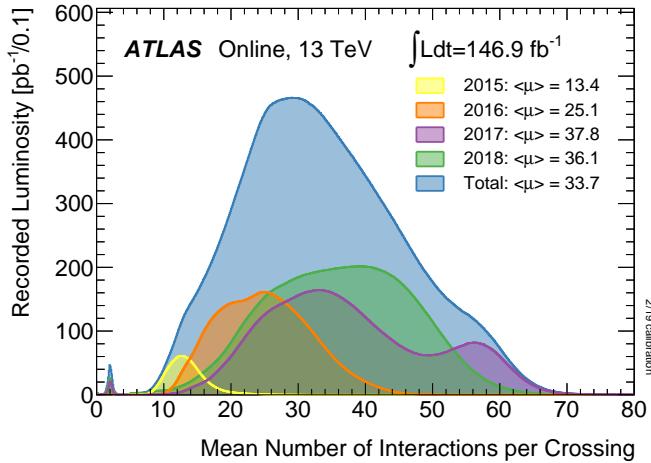


Figure 3.5: Average pile-up profiles measured by ATLAS during Run 2 [33]. Higher levels of pile-up are planned for Run 3.

420 3.3 The ATLAS Detector

421 The ATLAS² detector is made up of several specialised sub-detectors which are
422 arranged concentrically around the nominal interaction point at the centre of the
423 detector. The detector is designed to cover nearly the entire solid angle around the
424 collision point. In this section a condensed overview of each sub-detector is given, in
425 order of increasing radial distance from the point of collision. The inner tracking
426 detector is described in Section 3.3.1, the electromagnetic and hadronic calorimeters
427 in Section 3.3.2, the muon spectrometer in Section 3.3.3, and finally the trigger is
428 described in Section 3.3.4. More complete information on the detector can be found
429 in Ref. [31], while an overview of physics performance is given in [39].

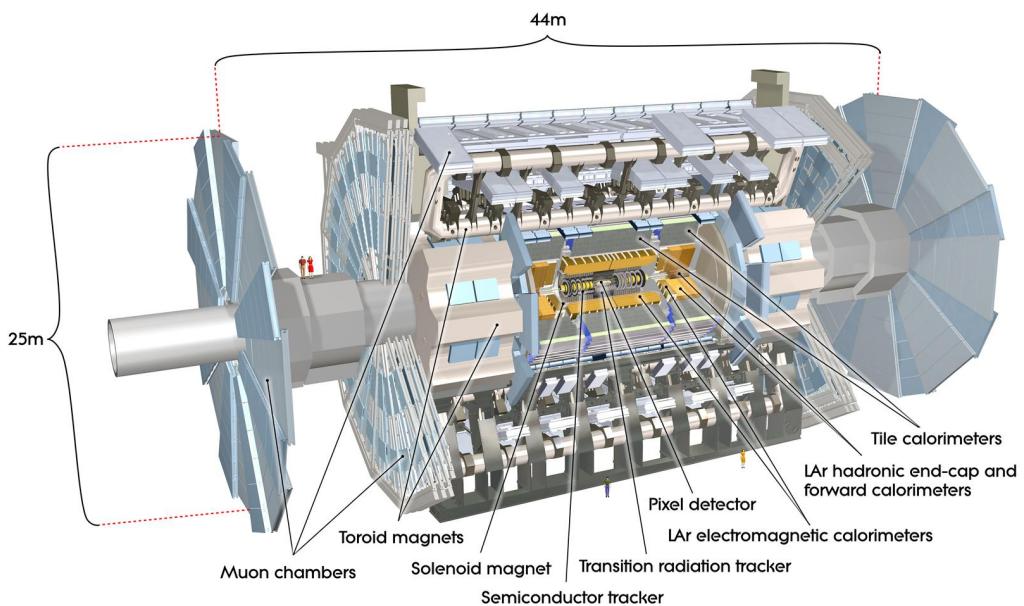


Figure 3.6: A 3D model of the entire ATLAS detector [40]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.

²A Toroidal LHC ApparatuS.

430 3.3.1 Inner Detector

431 The inner-detector system (ID) provides high-resolution charged particle trajectory
432 tracking in the range $|\eta| < 2.5$. The ID is immersed in a 2 T axial magnetic field,
433 produced by a superconducting solenoidal magnet, which enables the measurement
434 of particle momentum and charge. After Run 3, the ID will be replaced by the
435 ITk [41, 42].

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

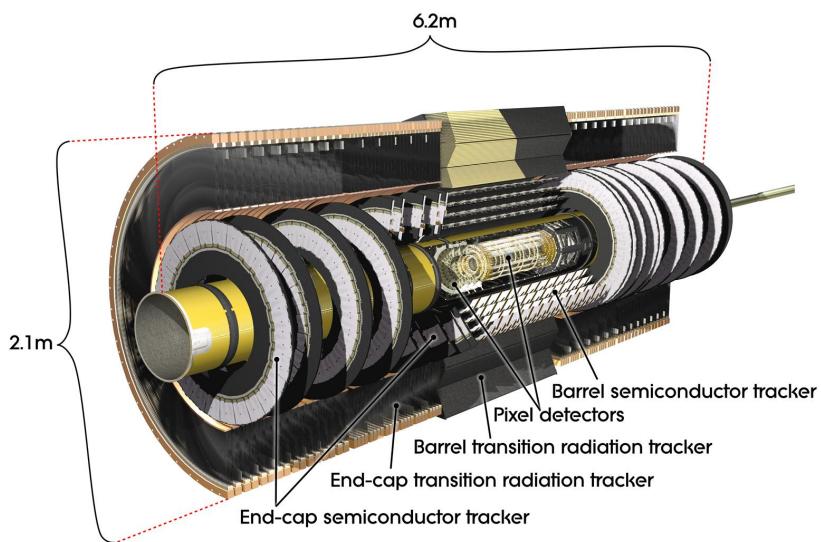


Figure 3.7: A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks [43].

442 The target inverse momentum resolution for the combined ID measurement is
443 parameterised as a function of the track transverse momentum and polar angle [39].
444 The parameterisation is given by

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

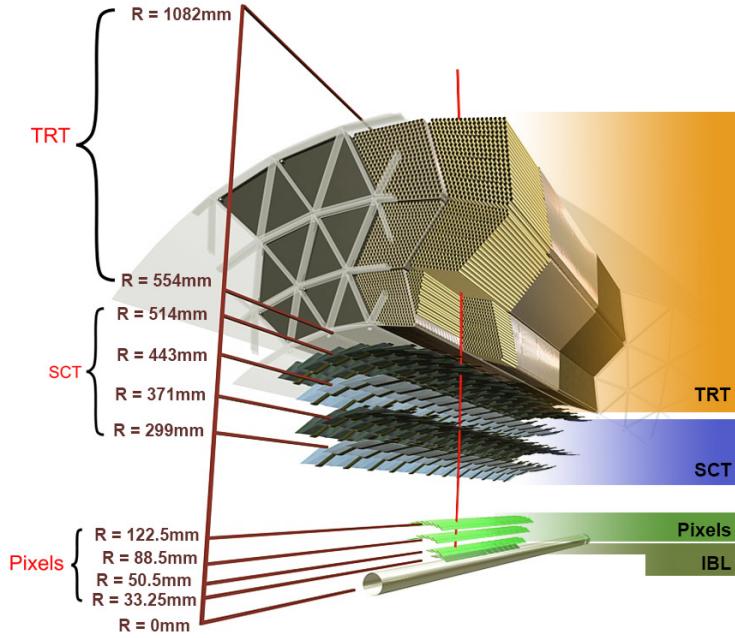


Figure 3.8: A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [36].

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

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

450 Pixel Detector

451 The silicon pixel detector is comprised of four cylindrical barrels at increasing radii
 452 from the beamline, and four disks on each side. The innermost barrel layer is
 453 the insertable B-layer (IBL), which was installed before Run 2 [44, 45] and lies
 454 approximately just 33 mm from the beam axis. The second-to-innermost layer is
 455 often referred to as the B-layer. The specification of the pixel detector determines the
 456 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 .

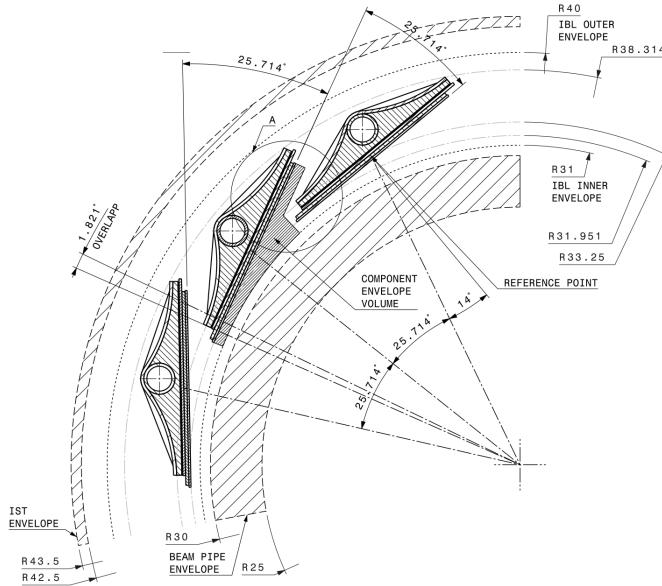


Figure 3.9: A schematic cross-sectional view of the ATLAS IBL [44].

462 Semi-Conductor Tracker (SCT)

463 The SCT is made up of four concentric barrel layers in the central region, and nine
 464 disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,
 465 with a small stereo angle (20 mrad) between the two layers enabling the z -coordinate
 466 to be measured from a pair of strip measurements. The SCT typically provides four
 467 precision spacepoint measurements (eight strip measurements) per track in the barrel
 468 region. These have intrinsic uncertainties of 17 μm in the transverse direction $R\phi$, and
 469 580 μm in the longitudinal direction z [46]. The measurements provide a contribution
 470 to the measurement of charged particle momentum and impact parameter, along
 471 with vertex position. Charge-particle tracks can be distinguished if separated by
 472 more than $\sim 200 \mu\text{m}$.

473 Transition Radiation Tracker (TRT)

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

487 3.3.2 Calorimeters

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

498 Liquid Argon (LAr) Electromagnetic Calorimeter

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

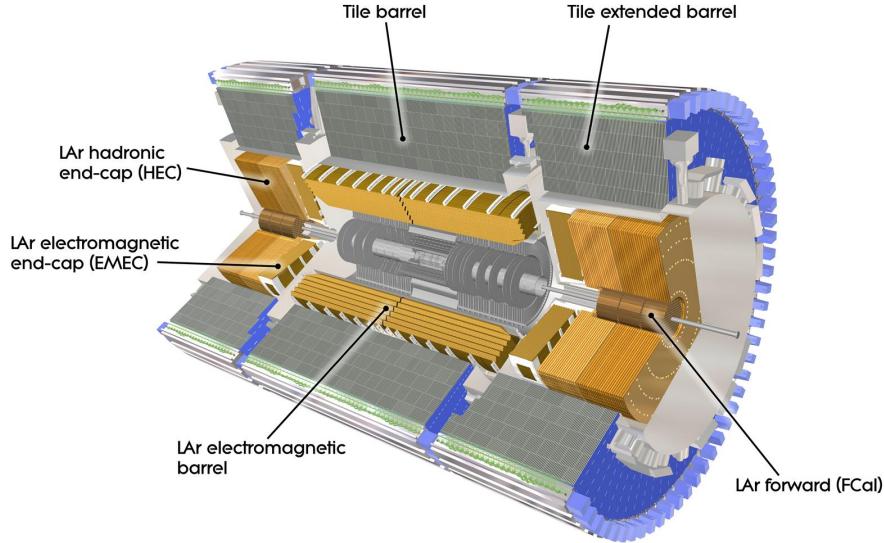


Figure 3.10: The ATLAS calorimeters [47]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic endcaps.

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

507 The energy resolution of the LAr calorimeter is made up of a sampling and a constant
 508 term, which are summed in quadrature to produce the overall energy resolution. The
 509 sampling term contributes approximately $10\%/\sqrt{E}$, while the constant term adds an
 510 additional 0.7%. Photons with moderate transverse energy $E_T \approx 50 \text{ GeV}$ have an
 511 overall energy resolution of at most 1.6% over most of the pseudorapidity range. At
 512 lower $E_T \approx 10 \text{ GeV}$, the resolution is degraded to approximately 5%. The resolution
 513 measurements are obtained from test beam data [39].

514 **Hadronic Tile Calorimeter**

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

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

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

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

523 corresponding to a jet energy resolution of 10% at a jet p_T of approximately 100 GeV
524 [48].

525 3.3.3 Muon Spectrometer

526 Due to their higher mass, muons easily pass unimpeded through the ID and calorime-
527 ters and therefore require specialised detectors for their measurement. The Muon
528 Spectrometer (MS) is made up of dedicated tracking and triggering hardware. The
529 precision tracking system uses three layers of monitored drift tubes with a barrel
530 region covering $|\eta| < 1.2$ and end-caps covering $1 < |\eta| < 2.7$. The inner layers of
531 the end-caps use cathode strip chambers to better cope with the high occupancy
532 in the forward region. Precision tracking resolution is approximately 50 μm . The
533 trigger system is comprised of resistive plate chambers in the barrel region covering
534 $|\eta| < 1.0$ and thin gap chambers in the end-cap regions covering $1 < |\eta| < 2.4$. A set
535 of three superconducting air-core toroidal magnets, each made up of eight coils, is
536 used in each of the barrel and end-caps to deflect the muons as they pass through
537 the MS, allowing their momentum and charge to be measured from the direction
538 and magnitude of curvature. The toroidal magnets generate a field which is largely
539 orthogonal to the muon trajectories which allows for maximum deflection. The
540 transverse momentum resolution has been measured to be approximately 1.7% in the
541 central region for low- p_T muons, increasing to 4% for high- p_T muons in the forward
542 regions [49].

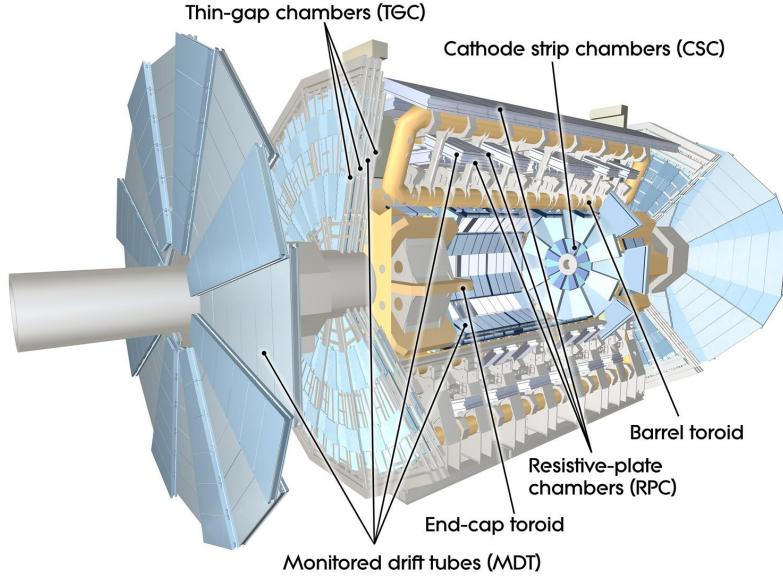


Figure 3.11: The ATLAS muon spectrometer [50].

543 3.3.4 The Trigger

544 The 25 ns bunch spacing used over the course of Run 2 corresponds to a bunch-
 545 crossing or event rate of 40 MHz (see Table 3.1). If the full information for the
 546 detector was written out for each event, this would correspond to the generation
 547 of 60 TB of data each second. This is more than can be feasibly read out from
 548 the hardware, processed and stored, requiring the use of a trigger system which
 549 quickly makes a decision about whether or not an event is potentially interesting
 550 and should be kept for further analysis. The trigger system is comprised of two
 551 levels which search for signs of electrons, muons, taus, photons, and jets, as well as
 552 events with large total or missing transverse energy. The hardware-based Level-1
 553 (L1) trigger uses coarse information from the calorimeters and MS to accept events at
 554 an average rate of 100 kHz approximately 2.5 μ s after the event. After the L1 trigger,
 555 the software-based High Level Trigger (HLT) makes use of 40 000 CPU cores to make
 556 a final selection on surviving events in approximately a few hundred milliseconds.
 557 The final event read-out rate is approximately 1.2 kHz, corresponding to 1.2 GB s^{-1}
 558 of permanent data storage. More information is provided in [51].

559 3.4 Reconstructed Physics Objects

560 Event reconstruction is the process of analysing the output from the detector to
561 determine the type and properties of particles present in an event. The reconstructed
562 event provides information about the underlying physics process that led to these
563 observable final state particles. Events passing the trigger selection (described in
564 Section 3.3.4) undergo offline reconstruction, which makes use of the full information
565 from the detector. Reconstruction and analysis of events relies on the extensive
566 ATLAS software stack, see Ref. [52] for more information.

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

569 3.4.1 Tracks

570 The reconstructed trajectories of charged particles are referred to as *tracks*. Track are
571 reconstructed from the energy depositions (called *hits*) left by the particles as they
572 traverse the the inner detector. Tracks are widely used for a variety of downstream
573 applications, including vertexing and jet tagging, so their accurate reconstruction
574 is a critical task. A comprehensive introduction to ATLAS tracking is available
575 in Ref. [53], while specific optimisations for dense environments are detailed in
576 Refs. [54, 55]. An overview of track reconstruction is given below.

577 Space-point Formation (Clustering)

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

584 Track Finding

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

601 Ambiguity Solving

602 Ambiguity solving was introduced as part of the ATLAS New Tracking effort [53],
603 which was intended to improve track reconstruction performance in dense envi-
604 ronments. In the ambiguity solver, track candidates are processed individually in
605 descending order of a track score. The track score quantifies the likelihood of the
606 track corresponding to the trajectory of a real particle. Scoring uses a number of
607 variables, including the number and positions of hits (preferring hits in more precise
608 regions of the detector), the transverse momentum of the track and the track fit
609 quality. The track fit quality describes the quality of the track as the χ^2 divided
610 by the number degrees of freedom on the track. A preference for high transverse
611 momentum tracks promotes the successful reconstruction of the more physically
612 interesting energetic particles, and suppresses the large number of wrong hits assigned
613 to low momentum tracks. The ambiguity solver also penalises tracks with missing
614 hits on the innermost detector layers.

615 During the processing of a given highest-scoring track candidate, the track is cleaned
616 (whereby problematic hits are removed), and, if the resulting track satisfies the quality
617 selection criteria, a high precision fit of the track parameters using the surviving hits
618 is performed. The high precision fit makes full use of all available information, and
619 uses an updated position and uncertainty estimate for each cluster obtained from
620 a Neural Network (NN) [56]. If the track has reached this stage without rejection
621 by passing various quality regiments, it is re-scored and returned to the list of track
622 candidates. If the same track is then processed again without requiring modification,
623 it is added to the final track collection. Track candidates that fall below a certain
624 quality cut are rejected. This selection does allow for the possibility of a track having
625 small number of shared hits.

list shared
hit cut?

626 Neural Network Cluster Splitting

627 As part of track cleaning, shared hits are classified by a NN to determine if they are
628 compatible with the characteristic features of a merged cluster [54, 56]. A merged
629 cluster is one made up of a combination of energy deposits from more than one
630 particle, which have become merged due to the closeness of the associated particles
631 and the limited resolution of the detector. While in general this event is rare, it
632 is common for clusters to become merged in dense environments, as discussed in
633 Section 4.1. If the cluster is predicted to be merged it is labelled as being freely
634 shareable, or *split*. Hits not compatible with the merged hypothesis can still be
635 shared by a limited number of tracks, but come with a penalty for the track which
636 may hinder its acceptance into the final track collection.

637 Pseudotracking

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

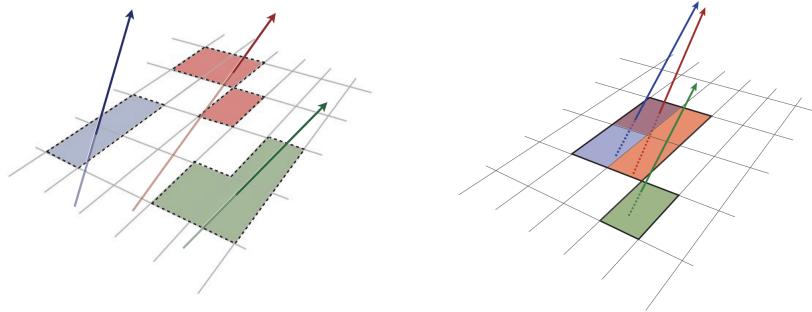


Figure 3.12: Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [55].

645 ATLAS detector, with perfect hit-to-track association and and track reconstruction
 646 efficiency. The approach was introduced in Ref. [57] as a way to obtain a fast
 647 approximation of tracking reconstruction for simulated data, however the technique
 648 has become a useful tool for studying tracking performance in general [54].

649 3.4.2 Vertices

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

658 Primary Vertices

659 Each proton-proton interaction happens at a *primary vertex* (PV). Primary vertices
 660 are iteratively reconstructed with tracks using the iterative vertex finder (IVF) [58].
 661 In Run 3, the IVF will be replaced with an adaptive multi-vertex finder (AMVF) [59].

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

664 **Secondary Vertices**

665 Secondary vertices (SV) occur when a particle radiates or decays at a sufficient
666 distance from the primary vertex to be resolved from the primary vertex (see
667 Section 4.1.1). Two widely used secondary vertexing tools are used within ATLAS:
668 SV1 and JetFitter [60]. Each attempts to reconstruct secondary vertices inside a jet
669 using the tracks associated to that jet (see Section 3.4.3 for more information about
670 track association). SV1 by design attempts to reconstruct only a single inclusive
671 vertex per jet. This inclusive vertex groups all b -hadron decay products, including
672 tracks from the b -hadron decay itself and tracks from $b \rightarrow c$ decays. The second tool,
673 JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary
674 vertices from b -hadron decays are reconstructed separately to tertiary vertices from
675 $b \rightarrow c$ decay chains.

676 **3.4.3 Jets**

677 Jets are an aggregate reconstructed object corresponding to a collection of collimated
678 stable particles which results from a decay chain of an quark or gluon progenitor. Jets
679 are built by clustering constituent objects (e.g. tracks or calorimeter clusters) using
680 a jet finding algorithm, for example the anti- k_t algorithm [61], which is implemented
681 in FASTJET [62].

682 **EMTopo Jets**

683 EMTopo jets are reconstructed from noise-suppressed topological clusters (topoclus-
684 ters) of calorimeter energy depositions. The clustering uses the energy significance
685 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. [63].

Particle Flow Jets

Particle-flow (PFlow) jets are reconstructed from particle-flow objects [64] using the anti- k_t algorithm with a radius parameter of 0.4. Particle-flow objects integrate information from both the ID and the calorimeters, improving the energy resolution at high transverse momenta and reducing pile-up contamination. The PFlow jet energy scale is calibrated according to Ref. [65].

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

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

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

Large- R Jets

Large- R jets have a radius parameter $R = 1.0$ and are built by clustering topological calorimeter clusters using the anti- k_t algorithm [66]. The large radius parameter is especially useful for containing the decay products of a boosted Higgs boson, as discussed in Chapter 7. Due to their large size, large- R jets benefit from a grooming procedure called trimming which remove soft contaminants inside the jet [67, 68].

715 Trimming aims to remove jet constituents from pile-up and the underlying event,
716 which helps to improve the jet mass resolution and its robustness to varying levels
717 of pile-up. The jet mass is computed using a combination of information from the
718 calorimeters and ID, and a calibration to data is applied [69].

719 **Track-jets**

720 Track-jets are built by clustering tracks using the anti- k_t clustering algorithm and
721 are used in the analysis described in Chapter 7. The radius parameter is allowed
722 to vary with transverse momentum such that a broader cone (up to $R = 0.4$) is
723 used for low- p_T track-jets and a narrower cone (down to $R = 0.02$) for high- p_T
724 track-jets [70, 71]. The narrower cone is better suited to clustering highly collimated
725 jet constituents at high- p_T . Truth flavour labels for track-jets are derived using the
726 same $\Delta R(\text{hadron}, \text{jet}) < 0.3$ matching scheme as used for PFlow jets.

727 **3.4.4 Leptons**

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

731 **Electrons**

732 Electrons candidates are reconstructed by matching PV-compatible³ inner detector
733 tracks to topological calorimeter clusters. The track-cluster matching criteria takes
734 into account the significant energy loss of the electron due to bremsstrahlung. If a
735 match is found, a refit of the track is performed using the Gaussian Sum Filter (GSF)
736 [72], which better handles trajectory reconstruction in the presence of bremsstrahlung.
737 Various identification criteria are then applied to the candidates using a likelihood-
738 based (LH) method to improve purity. These include requirements on the track
739 quality and cluster matching, the shape of electromagnetic shower in the ECal,
740 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.

- 741 Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may also be applied. A full description can be obtained from Ref. [73].

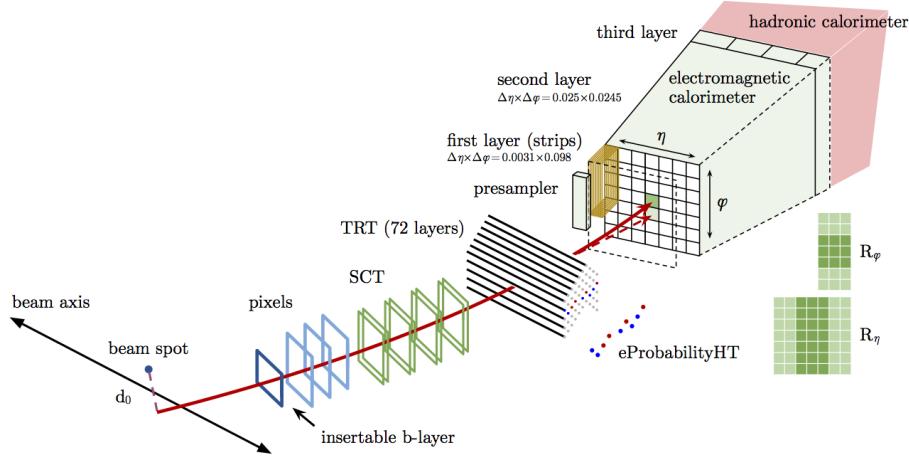


Figure 3.13: A sketch of electron reconstruction using the ATLAS detector [73]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.

742

743 Muons

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

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

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

765 More information on muon reconstruction, identification and isolation can be found
766 in Ref. [74].

767 3.4.5 Missing Transverse Momentum

768 An imbalance in the final state transverse momentum can occur as a result of
769 incomplete measurement of the final state particles. In particular, neutrinos are
770 not measured by the detector and contribute to the missing transverse momentum
771 $\mathbf{E}_T^{\text{miss}}$. Incomplete detector acceptance and inaccuracies in the reconstruction of the
772 final state can also contribute to the missing transverse momentum of an event. In
773 order to calculate the missing transverse momentum, the negative vector sum of
774 the momentum of all photons, leptons and small- R jets with $p_T > 20 \text{ GeV}$ is taken.
775 The momenta of tracks associated to the primary vertex are also taken into account.
776 The magnitude of $\mathbf{E}_T^{\text{miss}}$ is written E_T^{miss} . More information about missing transverse
777 momentum reconstruction is provided in [75].

778

Chapter 4

779

Tracking and b -tagging

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

788 Historically a two tiered approach to b -tagging has been taken, in which so called
789 *low-level* taggers take as inputs information about the jet and associated tracks, and
790 attempt to reconstruct or identify some aspect of a b -jet, such as displaced tracks
791 or secondary vertices. The outputs of several low-level taggers are then fed into a
792 *high-level* tagger, which uses a multivariate approach to discriminate between jet
793 flavours.

794 As the different b -tagging algorithms ultimately rely on tracks, accurate and efficient
795 track reconstruction is essential. This chapter summarises the challenges facing
796 tracking and b -tagging at high transverse momentum with an investigation into track
797 reconstruction performance in Section 4.1. Some preliminary investigations into
798 improving tracking in this regime are investigated in Section 4.2.

⁷⁹⁹ 4.1 *b*-hadron Reconstruction

⁸⁰⁰ This section outlines the typical detector signature of a *b*-hadron in Section 4.1.1
⁸⁰¹ and discusses some associated reconstruction difficulties in Section 4.1.2.

⁸⁰² 4.1.1 Decay Topology

⁸⁰³ *b*-hadrons are quasi-stable bound states of a bottom quark and one or more lighter
⁸⁰⁴ quarks. Collectively, these are the *B*-mesons (e.g. $B^+ = u\bar{b}$, $B^0 = d\bar{b}$) and baryons
⁸⁰⁵ (e.g. $\Lambda_b^0 = udb$). After a *b*-quark is produced as the result of a proton-proton collision,
⁸⁰⁶ they quickly hadronise. The hadronisation process is hard – around 70-80% of
⁸⁰⁷ the *b*-quark’s momentum is passed to the *b*-hadron, with the rest being radiated
⁸⁰⁸ as prompt hadronisation or fragmentation particles. See Ref. [76] for a more in
⁸⁰⁹ depth discussion on hadronisation and the closely related process of fragmentation.
⁸¹⁰ Henceforth the combined hadronisation and fragmentation products will be referred
⁸¹¹ to collectively as fragmentation.

⁸¹² *b*-hadrons are interesting objects of study due to their relatively long proper lifetimes
⁸¹³ $\tau \approx 1.5$ ps [77]. This lifetime corresponds to a proper decay length $c\tau \approx 450$ μm . In
⁸¹⁴ the rest frame of the detector, the typical *b*-hadron travels a distance

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

⁸¹⁵ before decaying, where in the high energy limit $\gamma = E_b/m_b$ and $\beta = v/c = 1$.

⁸¹⁶ For a 50 GeV *b*-hadron, this gives $d \approx 4.5$ mm, which is displaced enough to be
⁸¹⁷ resolved from the primary vertex. Meanwhile for a 1 TeV *b*-hadron, $d \approx 90$ mm –
⁸¹⁸ well beyond the radius of the first pixel layer (the IBL) which is situated at a radius
⁸¹⁹ of approximately 33 mm from the center of the detector (the distance varies due
⁸²⁰ to the interleaved structure) Fig. 4.1 shows how the mean decay radius varies as a
⁸²¹ function of *b*-hadron p_T . This significant displacement is characteristic of *b*-jets and
⁸²² makes it possible to reconstruct secondary vertices at the *b*-hadron decay point.

⁸²³ *b*-hadrons decay weakly to on average four or five collimated stable particles [78].
⁸²⁴ These particles, along with any other fragmentation particles, are reconstructed in
⁸²⁵ the detector as a jet. A *b*-jet has several characteristic features which differentiate

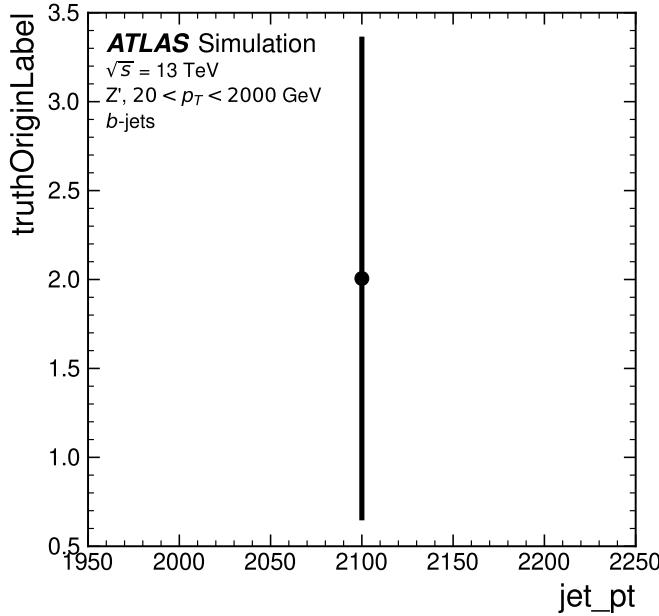


Figure 4.1: The truth *b*-hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed *b*-jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.

it from light-jets. These features stem from the significant displacement of the *b*-hadron that can occur due to its lifetime. The primary feature is the presence of a high mass secondary vertex that is significantly displaced from the primary vertex. Reconstruction of these vertices from tracks with common points of spatial origin is a common approach used in the identification of *b*-jets.

Additional signatures of *b*-hadrons are as follows. Associated tracks and SVs can have a large transverse impact parameter d_0 as a result of the *b*-hadron displacement (as shown in Fig. 4.2). Since it is common for the *b*-hadron to decay to a *c*-hadron with non-negligible lifetime, tertiary vertices can be found within *b*-jets resulting from $b \rightarrow c$ decay chains. The *b*-hadron also decays semileptonically in approximately 23% of cases [14]. The presence of a reconstructed electron or muon inside a jet can also be a key indicator that the jet was instantiated by a *b*-hadron.

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

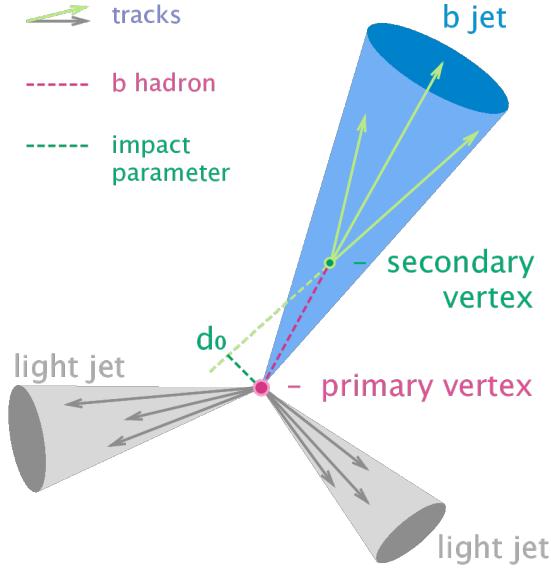


Figure 4.2: Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey) [79]. The b -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter d_0 is a characteristic property of the trajectories of b -hadron decay products.

842 4.1.2 Challenges Facing b -hadron Reconstruction

843 As discussed, a necessary requirement for successful b -tagging is the efficient and
 844 accurate reconstruction of the charged particle trajectories in the jet. For high p_T jets
 845 ($p_T > 200$ GeV) this task becomes difficult due to a combination of effects. As the
 846 b -hadron energy increases, the multiplicity of the fragmentation products inside the
 847 jet increases, while the multiplicity of the products of the weak decay is unaffected.
 848 The “signal” tracks (those from the weak decay of the b -hadron) therefore become
 849 outnumbered. Both fragmentation and b -hadron weak decay products also become
 850 increasingly collimated as their inherited transverse momentum increases. At high
 851 energies, the increased decay length of b -hadrons (and c -hadrons) means that decay
 852 products have less of an opportunity to diverge before reaching the first tracking
 853 layers of the detector (shown in Fig. 4.3). If the weak decay of the b -hadron takes
 854 place close enough to a detector layer, or if the particles are otherwise sufficiently
 855 collimated, charge deposits left by nearby particles may not be resolved individually,
 856 instead being reconstructed as merged clusters.

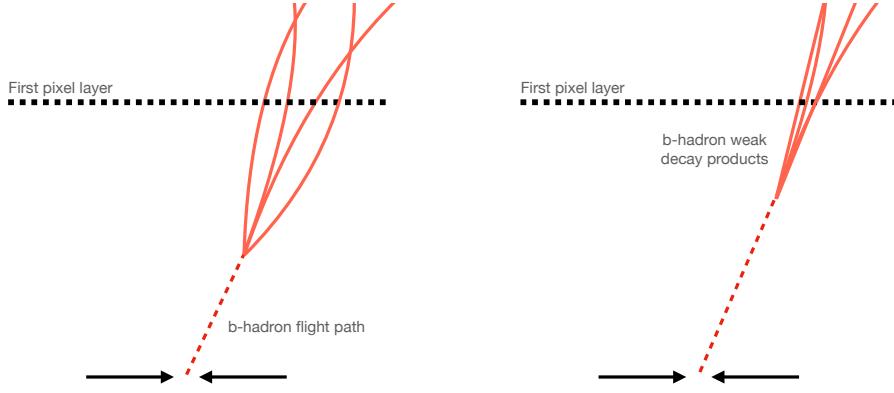


Figure 4.3: At lower p_T (left) the decay length of the b -hadron is reduced, and the resulting decay tracks are less collimated. At higher p_T (right) the b -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching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

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

868 The above effects create two related, but distinct problems for b -tagging. The first
 869 part is a drop in track reconstruction efficiency. The presence of shared and missing
 870 hits reduces a track's score in the ambiguity solver meaning that higher ranking, but
 871 potentially worse, track candidates are processed first and take ownership of the hits.
 872 This can make it difficult for otherwise reasonable b -hadron decay tracks to meet
 873 the ambiguity solver's stringent track quality requirements, leading to their rejection
 874 at this stage and an overall descrease in the b -hadron decay track reconstruction
 875 efficiency as shown in Fig. 4.4.

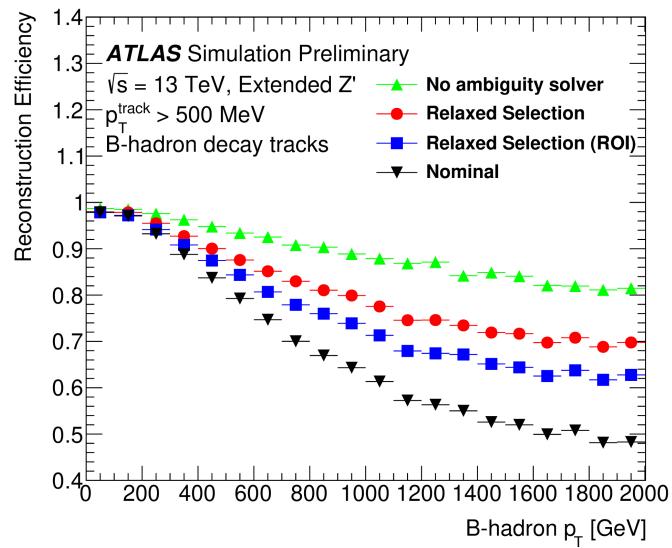


Figure 4.4: *b*-hadron decay track reconstruction efficiency as a function of truth *b*-hadron p_T [80]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- p_T *b*-hadrons, 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.

876 The second part of the problem is that, due to the high multiplicity of clusters available
 877 for assignment in the vicinity of the typical high energy b -hadron decay track, and
 878 also given the strong positive bias of the ambiguity solver towards those tracks
 879 with pixel measurements in each layer (especially the innermost IBL measurement),
 880 many b -hadron decay tracks are assigned incorrect inner layer hits. This is only a
 881 problem for those decay products which were produced within the pixel detector
 882 as a result of a significantly displaced b -hadron decay, and so do not have a correct
 883 hit available for assignment. Fig. 4.5 shows the number of hits as a function of the
 884 reconstructed track p_T for fragmentation tracks and tracks from the weak decay of
 885 the b -hadron. The baseline tracks represent the standard reconstruction setup, while
 886 the pseudotracks represent the ideal tracking setup as outlined in Section 3.4.1. The
 887 incorrect hits may skew the parameters of the track, which can in turn mislead the
 888 downstream b -tagging algorithms. In particular, b -tagging algorithms rely heavily on
 889 the transverse impact parameter significance $s(d_0)$ of the track. The quality of this
 890 measurement is expected to be adversely affected by wrong inner-layer hits on the
 891 track. Furthermore, multiple tracks sharing an incorrect hit can lead to the creation
 892 of spurious secondary vertices, which can cause further problems for the downstream
 893 b -tagging algorithms.

894 The combination of the effects described makes reconstructing tracks in the core
 895 of high p_T b -jets particularly challenging. The reduced reconstruction efficiency of
 896 b -hadron decay tracks and incorrectly assigned hits is thought to be the primary
 897 cause of the observed drop in b -tagging efficiency at high energies, however further
 898 study is required to determine which effect may dominate.

899 4.2 Investigations into High p_T b -hadron Tracking

900 In Section 4.2.1 pseudotracks, a key tool for studying the ideal tracking performance
 901 of the ATLAS detector, are used to study the shared hit requirements on tracks in
 902 the dense cores of high- p_T b -jets. Section 4.2.2 details a study which investigated
 903 modifying the global track fitter to improve reconstruction performance in this
 904 regime.

include plot
from sebs
study show-
ing they
are approx
similar im-
pacts? or
just mention
result? Can
do put need
to remove
ATLAS la-
bels. Alter-
natively you
can put an
internal refer-
ence to his
work and
state what
the outcome
is

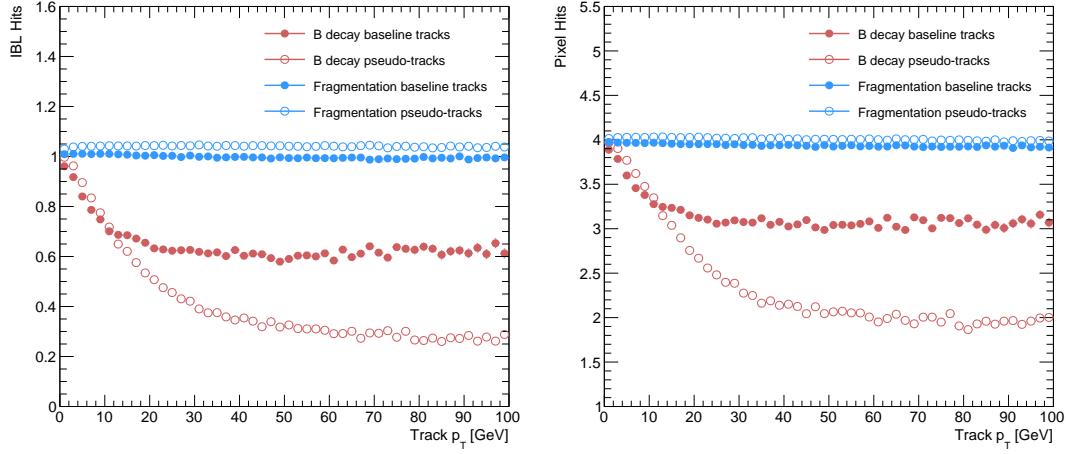


Figure 4.5: Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. Baseline tracks are those produced in the standard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

905 4.2.1 Shared Hits

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

921 that for the B-layer, the ambiguity solver is being overly aggressive in its rejection of
 922 shared hits.

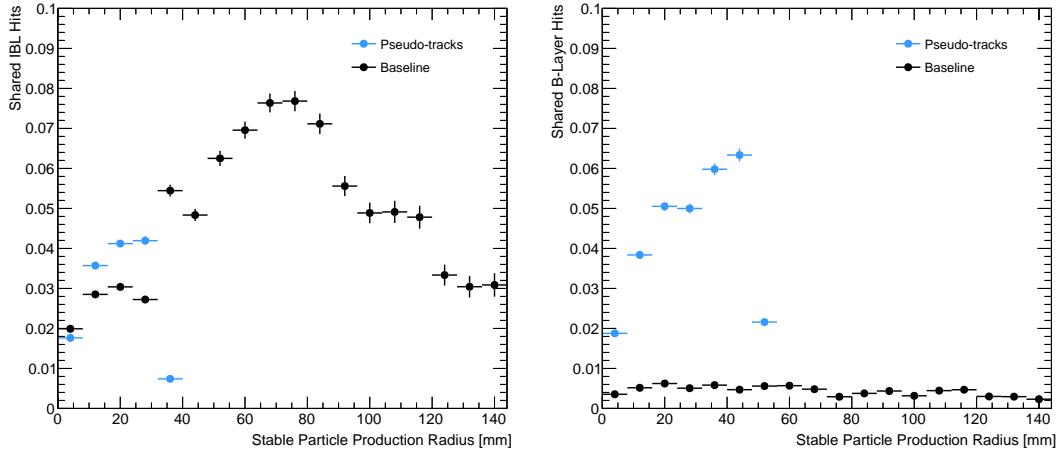


Figure 4.6: The fraction of hits which are shared on b -hadron decay tracks on the IBL (left), and the B-layer (right), as a function of the production radius of the b -hadron decay product. Pseudotracks represent the ideal performance given the ATLAS detector, see Section 3.4.1.

923 4.2.2 Global χ^2 Fitter Outlier Removal

924 This section documents ongoing progress into improvement of hit-to-track assignment
 925 by using the Global χ^2 Fitter (GX2F) to identify and prevent incorrect hits from
 926 being assigned to tracks during the track fit. This is in contrast to a previously
 927 investigated approach [81] which attempted to identify and remove wrong hits after
 928 the reconstruction of the track. As part of the track fit, an outlier removal procedure
 929 is run, in which suspicious hits are identified and removed.

930 The GX2F code, as a relatively low-level component of track reconstruction, has
 931 not undergone significant modification for several years, and was originally only
 932 optimised in the context of prompt, isolated tracks. During this time, a new tracking
 933 sub-detector, the IBL, was installed. The motivation for looking at the GX2F is that
 934 these changes may require re-optimisation of the GX2F code, and in particular the
 935 outlier removal procedures. Further motivation for this approach comes from the low
 936 rate of labelled outliers in baseline tracking. For example, while approximately 15%
 937 of b -hadron decay tracks have a wrong IBL hit (a value which only increases with

938 the p_T of the *b*-hadron), less than 1% of this tracks have had their IBL hit labelled
 939 and removed as an outlier.

940 **Implementation**

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

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

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

955 **Cut Optimisation**

956 A systematic variation of the outlier selection and χ^2/n thresholds has been carried
 957 out. Both thresholds were reduced in fixed step sizes of 0.25 for the outlier selection
 958 threshold and 1 for the χ^2/n threshold. The results for the best performing selections
 959 are discussed below. The value of the outlier selection threshold was reduced from 4
 960 down to 1.75, a change which affects all silicon layers (the TRT has separate outlier
 961 removal logic). Furthermore, a specific cut for the IBL was introduced, and is set
 962 to 1.25. The second threshold on the track χ^2/n was also reduced from 7 to 4.
 963 Finally, instead of taking the maximum of the pulls in the longitudinal and transverse

964 directions, a quadrature sum is taken of these two values and used. This variation is
 965 labelled “Mod GX2F” in plots.

966 The results are shown in Fig. 4.7 and demonstrate a reduction in wrong hit assignment
 967 whist also improving slightly the rate at which good hits are assigned to tracks. For a
 968 1 TeV track, the rate to assignment good hits to the track increases by approximately
 969 10%, while the rate to assign incorrect hits decreases by approximately 16%. The
 970 improvements are also observed when looking inclusively in all tracks, which avoids
 971 the need for a specific *b*-jet region-of-interest selection.

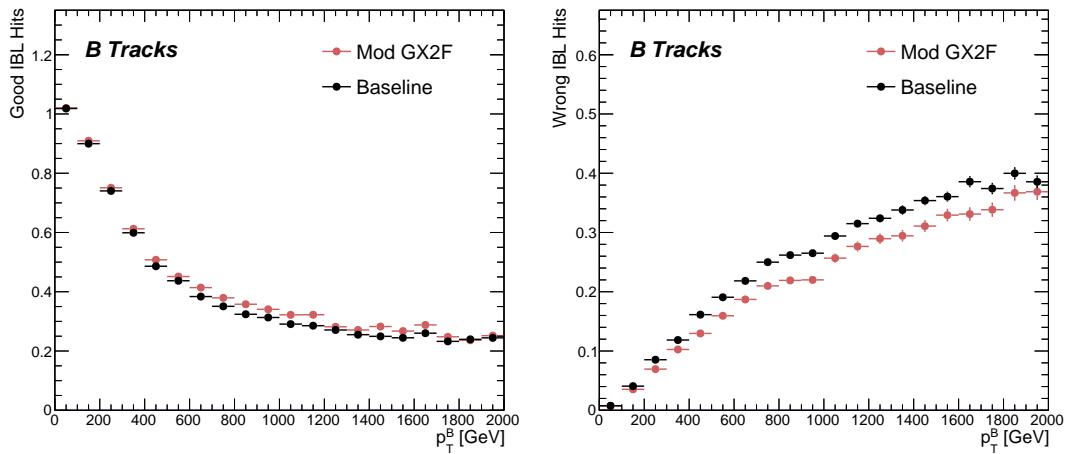


Figure 4.7: The rate to assign good (left) and wrong (right) IBL hits as a function of *b*-hadron p_T for tracks using baseline tracking (black) and the modified version of the outlier removal procedure (red). For each track, the corresponding p_T bin is filled with the number of good or wrong hits and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

972 An improvement, though modest, of all track parameter resolutions and pulls is
 973 observed. The improvement for the transverse impact parameter pull is shown in
 974 Fig. 4.8. The results demonstrate an improvement in hit assignment, unchanged
 975 reconstruction efficiency, and modest improvement in track parameter resolutions
 976 and pulls. In addition, the truth match probability of track is unchanged, suggesting
 977 that there is no increase in fake track rates. The changes are expected to have a
 978 negligible impact on computational resources.

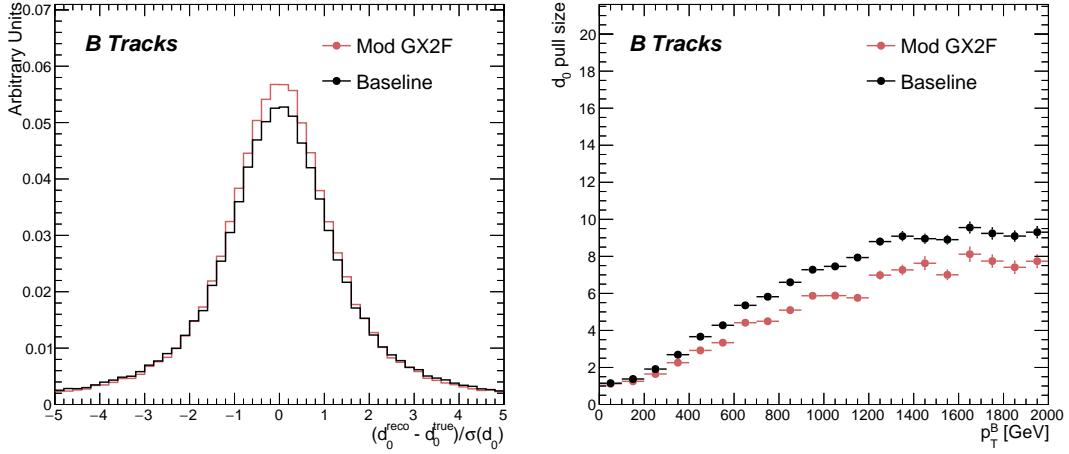


Figure 4.8: (left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified GX2F tracks. (right) The absolute value of the d_0 pull as a function of b -hadron transverse momentum.

4.3 Conclusion

In this section, the difficulties facing efficient and accurate track reconstruction, and hence performant b -tagging, have been outlined. The ambiguity solver, which attempts to clean or reject tracks which have an excessive number of shared hits, is shown to be overly aggressive in the removal of b -hadron decay product track candidates. The ambiguity solving process relies on a complicated pre-defined selection which has not been optimised for high transverse momentum b -hadron track reconstruction. These conclusions have motivated further ongoing studies into the improvement of the track reconstruction in dense environments and the high- p_T regime, such as those in Ref. [80].

An optimisation of the outlier removal process in the global χ^2 fitter was carried out. Though the results show some improvement over the baseline tracking scenario, these results need to be expanded upon by looking at the impact on the downstream b -tagging algorithms before putting them into production. As there are some known data-MC discrepancies, fine tuned optimisation such as the work presented here presents an opportunity to over-optimize the tracking algorithms on MC. The studies were carried out in Release 21 of the ATLAS software, and need to be reproduced using the newer Release 22 to confirm the results against other changes in the baseline tracking configuration. Thanks to the all-in-one flavour tagging approach described

998 in Chapter 6, it will also be easier in future to verify that the improvements to the
999 track reconstruction have a positive impact on the flavour tagging performance.

1000 **Chapter 5**

1001 **Track Classification MVA**

1002 The chapter details work on implementing a multivariate algorithm (MVA) to predict
1003 the truth origin of reconstructed tracks. An introduction to formalisms of machine
1004 learning is given in Section 5.1. In Section 5.2, the truth origin label is defined,
1005 and in Section 5.3 these labels are used to train a machine learning model that can
1006 effectively discriminate between good and fake tracks. Several studies motivated this
1007 work by demonstrating that at high p_T , b -tagging performance was degraded by the
1008 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm
1009 could be trained to detect fake tracks, these could be removed before their input to
1010 the b -tagging algorithms with the aim of improving performance.

1011 **5.1 Machine Learning Background**

1012 Over the past few decades, machine learning (ML) techniques have become increasing-
1013 ly popular in high energy physics experiments due the increased volumes of
1014 high-dimensional data and improvements in the techniques used (in particular deep
1015 learning). Machine learning is the process by which a computer program uses data
1016 to learn suitable parameters for a predictive model. This is opposed to explicitly
1017 providing instructions on how to perform a task. A subfield known as *supervised*
1018 *learning* is used in this work, and consists of exposing a model to a large number of
1019 labelled examples in order to extract relationships between the input data and their
1020 labels. These relationships are often complex, and explicitly programmed rules can
1021 fail to fully capture the relationships between inputs and outputs.

In the simplest case, a set of m labelled training examples $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$ is collected. Each element (x_i, y_i) consists of a input vector $x_i \in \mathbb{R}^{\text{input}}$, and the corresponding label y_i . In classification problems, these labels are integer *class labels* $y_i \in \{0, \dots, N - 1\}$, where N is the number of classes, which specify which of a pre-determined set of categorical classes the training example belongs to. The rest of the discussion in this chapter is limited to binary classification problems ($N = 2$). The two classes are often referred to as signal ($y_i = 1$) and background ($y_i = 0$), which need to be separated. Collecting sufficient and suitable data is one of the primary challenges of machine learning, as such data is not always readily available. Fortunately, sophisticated tools to simulate particle collisions have already been developed by the scientific community [82, 83]. These tools play a key role in generating a suitably large amount of labelled data which is used to train algorithms. More detail on the input datasets is given in Section 5.3.1.

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

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

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

5.1.1 Neural Networks

Neural networks (NNs) are a common choice for the machine learning model f since they have the ability to approximate any function [84] and are easy to train via backpropagation [85].

1051 **Artificial Neurons**

1052 The basic functional component of a NN is the *artificial neuron* or node, which is
1053 loosely inspired by a mathematical model of a biological neuron [86, 87]. A diagram
1054 of an artificial neuron is shown in Fig. 5.1 Each neuron is defined by its parameters
1055 or *weights* θ and a choice of activation function. Each neuron takes a fixed number
1056 of inputs and computes the dot product of the input and weight vectors $x^T \theta$ and
1057 additionally adds a constant bias term θ_0 . This term plays the role of a trainable
1058 constant value that is independent of the inputs.

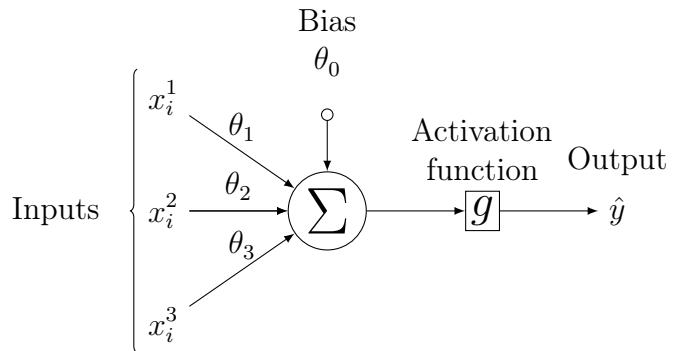


Figure 5.1: A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result z is passed to an activation function. Each neuron can be thought of as a logistic regression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

1059 The output of the dot product and bias term z is fed into an activation function
1060 g . The activation function has several uses, most notably acting as a source of
1061 non-linearity and bounding the output of the neuron. Some common activation
1062 functions (sigmod, tanh, ReLU and SiLU) are shown in Fig. 5.2. The choice of
1063 activation function can have implications for the performance and convergence of
1064 the network, since the gradient of g is used to compute the weight updates during
1065 training. This is also why input data is typically normalised to have zero mean and
1066 unity variance [88].

1067 **Networks**

1068 Several neurons are linked together in layers to form a neural network. The inputs
1069 are propagated layer-by-layer through the network until reaching the final output

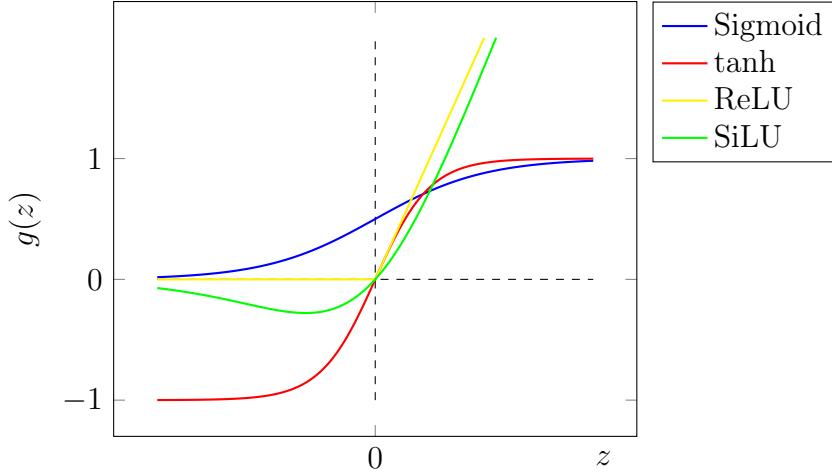


Figure 5.2: The output of several common choices for the activation function $g(z)$ of an artificial neuron. The input z is the output of the dot product between the activation and the weights, plus a bias term.

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

1085 problems, the binary cross entropy given in Eq. (5.3) loss if often used.

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

1086 Since the model f is differentiable, the error for each parameter θ_i can be computed by
 1087 taking the partial derivative of L with respect to the parameter. Updated parameters
 1088 θ'_i are calculated by updating the original parameter in the direction which reduces
 1089 the loss.

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

1090 The hyperparameter α is known as the *learning rate* and dictates the size of the
 1091 step taken in the direction of the slope. The errors for each parameter are efficiently
 1092 calculated using the backpropagation algorithm [85]. The process of updating weights
 1093 is repeated until the weights converge, which means the network is trained. In practice,
 1094 small batches of the input data are shown to the network at a time. For each batch
 1095 the average loss is calculated and the network's weights are updated. There are many
 1096 extensions and variations of the gradient descent algorithm. This work uses the Adam
 1097 optimiser which adds momentum to the weight updates (dampening oscillations)
 1098 and an adaptive per-parameter learning rate [89].

1099 5.2 Track Truth Origin Labelling

1100 Crucial to supervised learning techniques are the ground truth class labels which the
 1101 machine learning model is trained to predict. A set of track truth labels which a
 1102 high degree of granularity have been implemented in the ATLAS software stack, and
 1103 are listed in Table 5.1. The labelling scheme has been designed to be useful beyond
 1104 the classification of good and fake tracks. The origins are determined by analysing
 1105 the simulated record to determine the physical process that led to the creation of
 1106 the truth (i.e. simulated) particle which is associated with each reconstructed track.
 1107 Tracks are associated with truth particles by selecting the particle with the highest
 1108 *truth-matching probability* (TMP), defined in Eq. (5.5). This is a weighted sum of
 1109 the number of hits on a reconstructed track which are from the same truth particle,
 1110 versus the total number of hits on the track. The weights are subdetector-dependent
 1111 and are designed to account for the varying importance of the different subdetectors

1112 (based upon their precision) in the reconstruction of a track.

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

1113 For the fake track classification tool, the track truth origins in Table 5.1 are used
1114 to construct a binary label by assigning all fake tracks to the background category,
1115 and all other tracks as signal. The fake track classifier is then trained to distinguish
1116 between these two categories of tracks. Fake tracks are defined using the TMP,
1117 with a $\text{TMP} < 0.75$ ¹ giving a track the label of fake. Fake tracks are made up of
1118 combinatorial fakes, which are tracks which do not correspond to the trajectory of
1119 any truth particle, and poorly reconstructed tracks, which may somewhat resemble
1120 the trajectory of a truth particle due to the presence of some wrong hits on the track,
1121 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 [55]. 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.

1122 5.3 Fake Track Identification Tool

1123 The rate of fake tracks increases at high transverse momentum as shown in Fig. 5.3 due
1124 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-efficiency decreases by up to 35% at a b -efficiency of 35%.

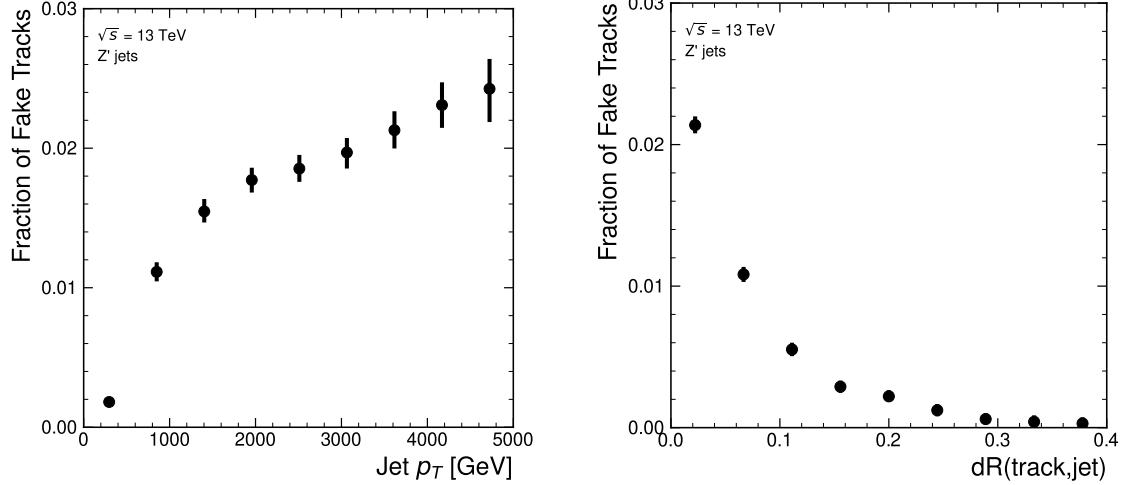


Figure 5.3: Rate of fake tracks as a function of jet transverse momentum (left) and $\Delta R(\text{track}, \text{jet})$ (right). The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases.

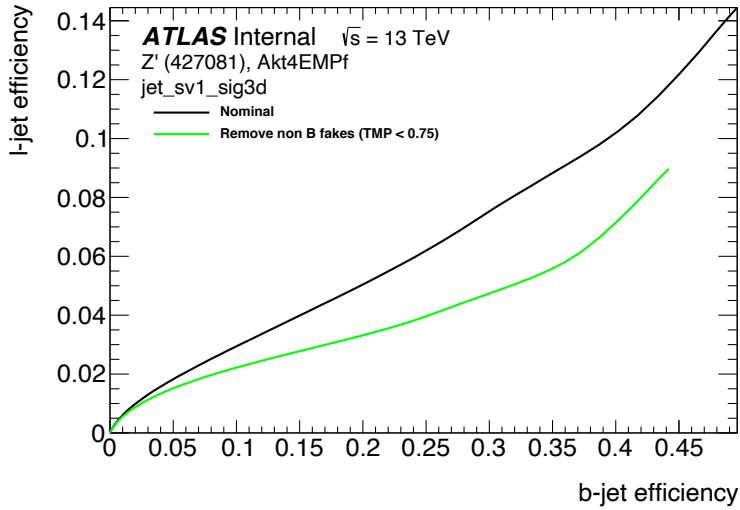


Figure 5.4: The light-jet efficiency of the low level tagger SV1 as a function of b -jet efficiency for the nominal tracking setup (black) and for the case where fake tracks which are not from the decay of a b -hadron are removed. The light-jet efficiency is decreased, demonstrating that the presence of fake tracks is detrimental to the algorithm performance.

1128 To identify and remove fake tracks, a NN classification tool was trained with all
1129 non-fake tracks as the signal class and fake tracks as the background class. Inputs to
1130 the model are described in Section 5.3.2, while fake track removal performance is
1131 given in Section 5.3.4.

1132 **5.3.1 Datasets**

1133 To train and evaluate the model, simulated SM $t\bar{t}$ and BSM Z' events initiated by
1134 proton-proton collisions at a center of mass energy $\sqrt{s} = 13$ TeV are used. The Z'
1135 sample is constructed in such a manner that it has a relatively flat jet p_T spectrum
1136 up to 5 TeV and decays democratically to equal numbers of b -, c - and light-jets.
1137 The generation of the simulated event samples includes the effect of multiple pp
1138 interactions per bunch crossing with an average pileup of $\langle \mu \rangle = 40$, which includes
1139 the effect on the detector response due to interactions from bunch crossings before
1140 or after the one containing the hard interaction.

1141 The $t\bar{t}$ events are generated using the POWHEGBox [90–93] v2 generator at next-
1142 to-leading order with the NNPDF3.0NLO [94] set of parton distribution functions
1143 (PDFs). The h_{damp} parameter² is set to 1.5 times the mass of the top-quark (m_{top}) [95],
1144 with $m_{\text{top}} = 172.5$ GeV. The events are interfaced to PYTHIA 8.230 [96] to model the
1145 parton shower, hadronisation, and underlying event, with parameters set according
1146 to the A14 tune [97] and using the NNPDF2.3LO set of PDFs [98]. Z' events are
1147 generated with PYTHIA 8.2.12 with the same tune and PDF set. The decays of b -
1148 and c -hadrons are performed by EVTGEN v1.6.0 [99]. Particles are passed through
1149 the ATLAS detector simulation [100] based on GEANT4 [101].

1150 Jets are required to have a pseudorapidity $|\eta| < 2.5$ and $p_T > 20$ GeV. Additionally, a
1151 standard selection using the Jet Vertex Tagger (JVT) algorithm at the tight working
1152 point is applied to jets with $p_T < 60$ GeV and $|\eta| < 2.4$ in order to suppress pile-up
1153 contamination [102].

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.

5.3.2 Model Inputs

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

Track selection follows the loose selection described in Ref. [103] and outlined in Table 5.3, which was found to improve the performance compared to previous tighter selections, whilst ensuring good resolution of tracks and a low fake rate [55]. Inputs are scaled to have a central value of zero and a variance of unity before training and evaluation.

5.3.3 Model Hyperparameters

Due to the imbalance between the two classes (with fake tracks being relatively uncommon), a weight was added to the loss function for the background class to account for this. The NN was made up of two hidden layers with 220 nodes per layer. The ReLU activation function was used in conjunction with the Adam optimiser with a learning rate of $1e-3$. Optimisation of the networks architecture was carried out to ensure optimal performance with a relatively small number of learnable parameters – 54 thousand. The model was trained using 40 million tracks with a futher 1 million tracks each used for validation and testing. A full list of the model hyperparameters 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 [55], 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 [55]. 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.

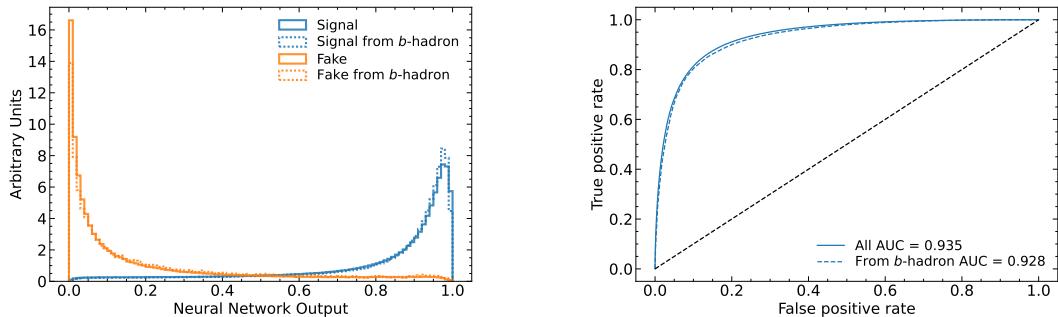


Figure 5.5: (left) Normalised histogram of the model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed line). The model is able 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 -efficiency of 70%, the light-jet mistag rate for jets with $250 < p_T < 400$ GeV is

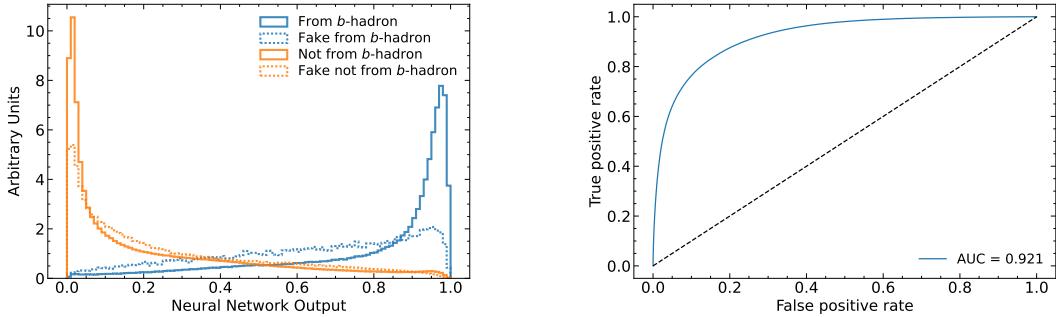


Figure 5.6: (left) Normalised histogram of the model output separated for tracks from the decay of a b -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line).

WP	Fake MVA Cut	b -hadron MVA Cut	Decay	Retained b -hadron Tracks	Fake & Non b -hadron Tracks Rejected
A	0.5	0.4		98.6%	50.7%
B	0.6	0.5		97.5%	62.0%

Table 5.6: 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).

reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For jets with $400 < p_T < 1000$ GeV the mistage rate drops from 0.1 to 0.08 for a similar relative improvement of 20%. The performance of the track classification tools is also shown for JetFitter in Fig. 5.8. The results are generally consistent with those from SV1, with up to a 20% improvement in the light-jet rejection for a given b -efficiency WP in the two jet p_T bins.

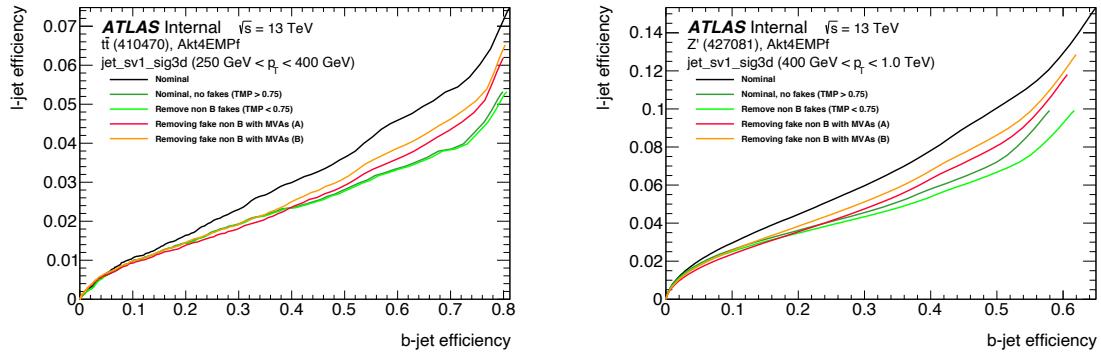


Figure 5.7: The effect of applying the fake track identification algorithm alongside the b -hadron decay track identification on the jet tagging performance of SV1 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet 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. It was found that b -hadron decay tracks can also be poorly reconstructed and thus marked as fake, it was necessary also to train a second algorithm to detect b -hadron decay tracks so that the removal of these tracks could be avoided. Removing fake and non- b decay tracks in this way was found to improve the light-jet mistagging rate of SV1 and JetFitter by up to 20% at high transverse momentum. The improved achieved using the classification tools was in general comparable with

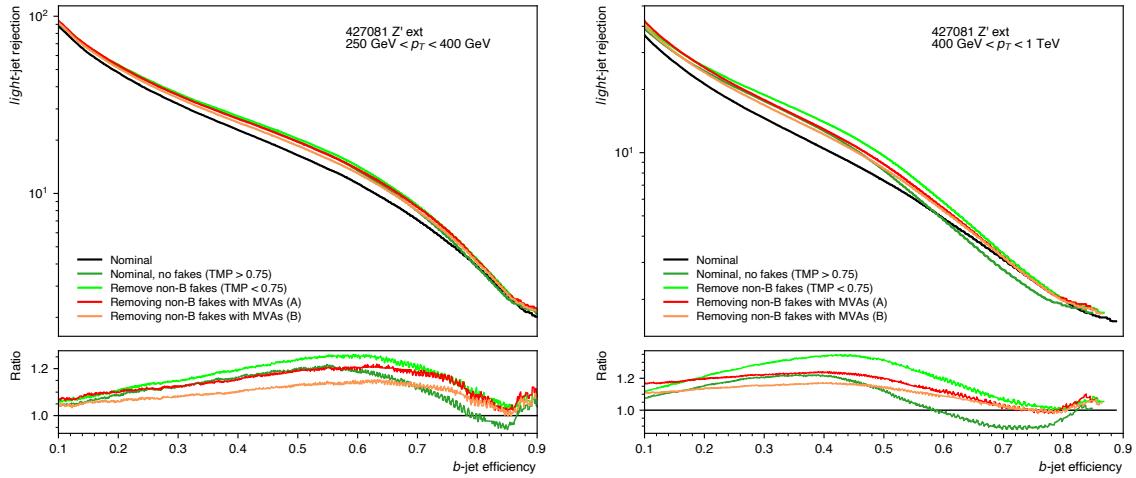


Figure 5.8: The effect of applying the fake track identification algorithm alongside the b -hadron decay track identification on the jet tagging performance of JetFitter for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal JetFitter light-jet rejection (black) is compared to two working points of fake track removal, labelled “A” (red) and “B” (orange). Removal of fake tracks based on truth information is shown by the green curves.

1239 the improvement achieved when using the truth information to remove the fake
1240 tracks not from the decay of a b -hadron.

1241 **Future Work**

1242 While removing tracks prior to their input to the low level tagging algorithms is
1243 shown here to be beneficial, a more performant alternative might be to keep these
1244 tracks but label them as being fake (for example using the output of the classification
1245 tool), and allow the tagging algorithms to take this into consideration, potentially
1246 making use of this information. This is not straightforward with manually optimised
1247 taggers such as SV1 and JetFitter, but is possible with more advanced taggers as
1248 described in Chapter 6.

1249 Tools which identify the origin of a given track have other potential uses. One
1250 application is to isolate a relatively pure sample of fake tracks which can be used
1251 to estimate the fake track rate in data, which would be useful for estimating the
1252 uncertainty on fake track modelling. Another application would be to use the
1253 b -hadron track identification tool to improve the track-to-jet association. Both
1254 applications are currently under investigation.

1255 The approach here works on a track-by-track basis, but a more sophisticated approach
1256 would consider the correlations between the tracks inside a jet, as shown in Chapter 6.

1257 Also left for future work is to simultaneously train a single tool which discriminates
1258 between all the truth origins listed in Table 5.1. Such a tool would be useful as a
1259 general purpose multiclass classifier.

₁₂₆₀ **Chapter 6**

₁₂₆₁ **Graph Neural Network Flavour
Tagger**

₁₂₆₃ Some of the work in this chapter has previously been published in Ref. [104]. The
₁₂₆₄ author of this thesis was on the editorial team and was the primary author. As such,
₁₂₆₅ figures and tables from the published note are reproduced here.

₁₂₆₆ As discussed in Chapter 4, flavour tagging is the identification of jets originating from
₁₂₆₇ b - and c -quarks. Flavour tagging is a critical component of the physics programme of
₁₂₆₈ the ATLAS experiment. It is crucial importance for the study of the Standard Model
₁₂₆₉ (SM) Higgs boson and the top quark, which decay preferentially to b -quarks [105, 106],
₁₂₇₀ and additionally for several Beyond Standard Model (BSM) resonances that readily
₁₂₇₁ decay to heavy flavour quarks [107].

₁₂₇₂ Existing flavour tagging algorithms, such as DL1r [108, 109], use as inputs the outputs
₁₂₇₃ from a number of low-level algorithms. The low-level algorithms each reconstruct
₁₂₇₄ various features of jets using the tracks which have been associated to the jet. The
₁₂₇₅ outputs from the low-level algorithms are then combined in a machine learning model.
₁₂₇₆ This chapter introduces GN1, a novel ML-based flavour tagging algorithm based on
₁₂₇₇ graph neural networks (GNNs).

₁₂₇₈ As opposed to the previous two-tiered approach described in Chapter 4, which requires
₁₂₇₉ the use of both low- and high-level algorithms, GN1 takes as inputs information from
₁₂₈₀ an unordered variable number of tracks to predict the jet flavour without requiring
₁₂₈₁ outputs from the intermediate low-level algorithms. In addition to predicting the
₁₂₈₂ flavour of the jet, the model predicts which physical processes produced the various

1283 jet tracks, and groups the tracks in the jet into vertices. These auxiliary training
1284 objectives provide valuable additional information about the contents of the jet and
1285 enhance the performance of the primary flavour prediction task.

1286 GN1 outperforms the existing ATLAS flavour tagging algorithms as shown in Sec-
1287 tions 6.5.1 and 6.5.2. For a b -jet efficiency of 70%, the light (c)-jet rejection is
1288 improved by a factor of ~ 1.8 (~ 2.1) for jets coming from $t\bar{t}$ decays with transverse
1289 momentum $20 < p_T < 250$ GeV. For jets coming from Z' decays with transverse
1290 momentum $250 < p_T < 5000$ GeV, the light (c)-jet rejection improves by a factor ~ 6
1291 (~ 2.8) for a comparative 30% b -jet efficiency.

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 [110, 111].

1302 The current ATLAS flavour tagger, DL1r [108], is a deep neural network which takes
1303 the outputs of a number of independently optimised “low-level” algorithms [60] as
1304 its inputs. Each of these low-level algorithms reconstructs a distinct feature of the
1305 experimental signature of heavy flavour jets using the tracks associated to the jet,
1306 and outputs some variables which are discriminating in the jet flavour.

1307 The low-level algorithms are a combination of manually optimised reconstruction
1308 algorithms, for example the SV1 and JetFitter algorithms that reconstruct displaced
1309 decay vertices, and trained taggers such as RNNIP and DIPS that use the IP and
1310 hit information from a variable number of tracks to identify the flavour of the
1311 jet [60, 103, 112, 113]. In contrast GN1 consists of only a single neural network, which
1312 takes the tracks as inputs along with some kinematic information about the jet. As a

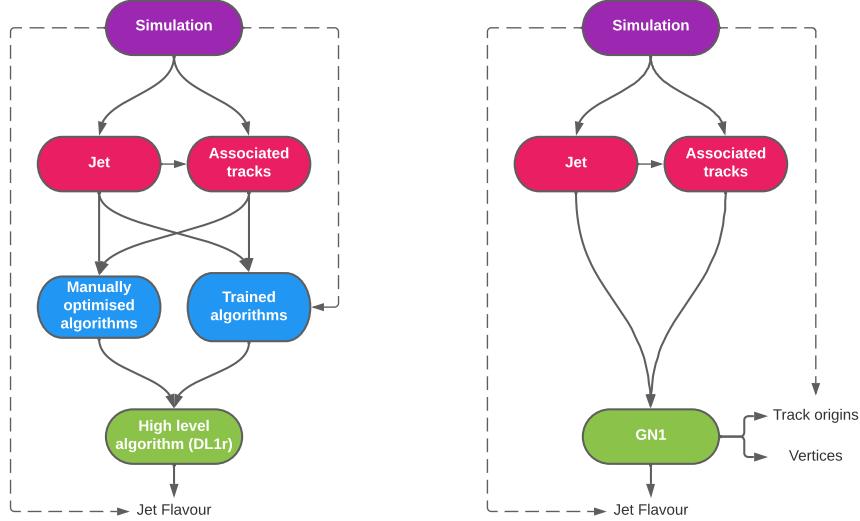


Figure 6.1: Comparison of the existing flavour tagging scheme (left) and GN1 (right). The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, additional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

result, it does not depend on the outputs of any other flavour tagging algorithm. A simple training of the model fully optimises its parameters, representing a significant simplification with respect to the optimisation procedure for DL1r. This is particularly important when optimising the tagger for new regions of phase space (e.g. c -tagging or high- p_T b -tagging), or when the detector or charged particle reconstruction algorithms are updated.

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

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

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

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. [55].

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 this section following the formalism in Ref. [114].

A graph \mathcal{G} consists of a set of N^n nodes $\mathcal{N} = \{h_i\}_{i=1:N^n}$ (also called *vertices*), a set of N^e edges $\mathcal{E} = \{e_i\}_{i=1:N^e}$, and a global representation u . Each node represents a 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, one for each of the nodes, edges, and global graph representation, and three aggregation functions. 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 using the global network ϕ^u as in

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

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

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

1366 For example, the global graph representation need not be present, and it is also
 1367 possible that no dedicated edge features are present. In such cases the corresponding
 1368 update and aggregation functions are not needed. Fig. 6.2 shows two possible graph
 1369 network update layers.

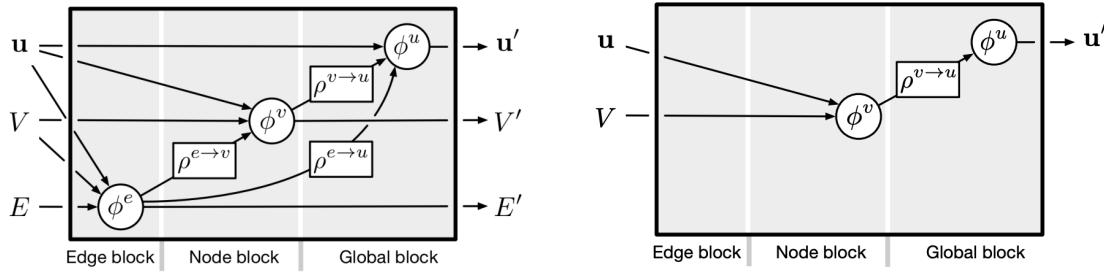


Figure 6.2: The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [115]. Reproduced from Ref. [114].

1370 The layer used in the GN1 model is specified in more detail in Section 6.4.3.

1371 6.3 Experimental Setup

1372 6.3.1 Datasets

1373 Datasets used to train the GN1 tagger are the same as described in Section 5.3.1.

1374 The training dataset contains 30 million jets, 60% of which are $t\bar{t}$ jets and 40% of
 1375 which are Z' jets. Although DL1r uses 70% $t\bar{t}$ jets and 30% Z' jets, the composition
 1376 of the training samples was found to have a negligible impact on the final performance
 1377 of GN1

1378 In order to evaluate the performance of the model, a statistically independent set of
 1379 500k testing jets from both the $t\bar{t}$ and Z' were used. Before being fed into the model,
 1380 the track- and jet-level inputs were normalised to have a mean of zero and a variance
 1381 of unity.

1382 Jet flavour labels are assigned according to the presence of a truth hadron within
 1383 $\Delta R(\text{hadron}, \text{jet}) < 0.3$ of the jet axis. If a b -hadron is found the jet is labelled a
 1384 b -jet. In the absence of a b -hadron, if a c -hadron is found the jet is called a c -jet.

1385 If no b - or c -hadrons are found, but a τ is found in the jet, it is labelled as a τ -jet,
1386 else it is labelled as a light-jet. Truth labelled b -, c - and light-jets are kinematically
1387 re-sampled in p_T and η to ensure identical distributions in these variables.

1388 6.4 Model Architecture

1389 6.4.1 Model Inputs

1390 As inputs, GN1 is fed two kinematic jet variables and an unordered set of up to 40
1391 tracks which have been associated to the jet. Each track consists of 21 variables. The
1392 kinematic jet variables are the jet transverse momentum and signed pseudorapidity.
1393 The input variables which are provided for each track are listed in Table 6.1. The
1394 track parameters and associated uncertainties, along with detailed hit information,
1395 are used as these variables carry critical information about the jet flavour. If more
1396 than 40 tracks are associated to a given jet, only the first 40 tracks with the largest
1397 transverse IP significance² $s(d_0)$ are fed into the model as inputs.

1398 In cores of high- p_T jets, tracks density is high due to the increased multiplicity and
1399 collimation fo tracks (see Chapter 4). As a result, the separation between tracks can
1400 be of the same order as the active sensor dimensions, resulting in merged clusters
1401 and tracks which share hits [55]. Due to the relatively long lifetimes of b -hadrons
1402 and c -hadrons, which can traverse several layers of the ID before decaying and have
1403 highly collimated decay products, the presence of shared or missing hits is a critical
1404 signature of heavy flavour jets.

1405 Dependence of the model on the absolute value of the azimuthal jet angle ϕ is
1406 explicitly removed by providing only the azimuthal angle of tracks relative to the jet
1407 axis. The track pseudorapidity is also provided relative to the jet axis. The sign of
1408 the jet pseudorapidity is included, but could be removed in the future to also build
1409 in the forward-backwards symmetry present at ATLAS.

1410 Since heavy flavour hadrons can decay semileptonically, the presence of a recon-
1411 structed lepton in the jet carries discriminating information about the jet flavour. To

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 [37].

1412 exploit this, a variant of GN1 called GN1Lep is trained in addition to the baseline
1413 model. The GN1Lep variant is identical to the baseline model, up to the inclusion
1414 an additional track-level input, leptonID, which indicates if the track was used in
1415 the reconstruction of an electron, a muon or neither. The variable is signed by
1416 the charge of the reconstructed lepton. The leptons used in the definition of the
1417 leptonID variable are required to satisfy basic quality requirements. The muons are
1418 required to be combined [116], and the electrons are required to pass the *VeryLoose*
1419 likelihood-based identification working point [117].

1420 The selections applied to the tracks is the same as that used for the fake track
1421 classification MVA described in Chapter 5. The full set of track selections is listed
1422 in Table 5.3. This selection was found to improve the flavour tagging performance
1423 compared to previous tighter selections, whilst ensuring good resolution of tracks
1424 and a low fake rate [55]. However, Section 6.5.8 demonstrates that further relaxation
1425 of the track selection requirements may be warranted.

1426 6.4.2 Auxiliary Training Objectives

1427 In addition to the jet flavour classification, two auxiliary training objectives are
1428 defined. The first auxiliary objective is the prediction of the origin of each track
1429 within the jet, while the second is the prediction of track-pair vertex compatibility.
1430 Each auxiliary training objective comes with a training target which, similar to
1431 the jet flavour label, is a truth labels derived from the simulation. The presence
1432 of the auxiliary training objectives improves the jet classification performance as
1433 demonstrated in Section 6.5.4.

1434 For the track origin prediction objective, each track is labelled with one of the
1435 exclusive categories defined in Table 5.1 after analysing the particle interaction that
1436 led to its formation. Since the presence of different track origins is strongly related to
1437 the flavour of the jet, training GN1 to recognise the origin of the tracks provides an
1438 additional handle on the classification of the jet flavour. This task may also aid the
1439 jet flavour prediction by acting as a form of supervised attention [118] - in detecting
1440 tracks from heavy flavour decays the model may learn to pay more attention to these
1441 tracks.

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

Table 6.1: Input features to the GN1 model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [55], 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.

1442 The vertexing auxiliary objective makes use of the fact that displaced decays of b -
 1443 and c -hadrons lead to secondary and tertiary vertices inside the jet, as described
 1444 in Section 4.1.1. The presence of displaced secondary vertices is not a completely
 1445 clean signal of a heavy flavour jet, as displaced secondary vertices can also occur in
 1446 light-jets as a result of material interactions and long-lived particle decays (e.g. K_S^0
 1447 and Λ^0). For the auxiliary object, GN1 predicts a binary label for each pair of tracks
 1448 in the jet. The label has a value of 1 if the truth particles associated with the two
 1449 tracks in the pair originated from the same spatial point, and 0 otherwise. To derive
 1450 the corresponding truth labels for training, truth production vertices within 0.1 mm
 1451 are merged, as these are assumed to be unresolvable given the granularity of the ID.
 1452 Track-pairs where one or both of the tracks in the pair have an origin label of either
 1453 Pileup or Fake are given a label of 0. Using the pairwise predictions from the model,
 1454 groups of tracks that have common compatibility can be formed, resulting in the
 1455 finding of vertices. Two existing low-level tagging algorithms, SV1 and JetFitter, are
 1456 currently used to find and reconstruct vertices inside jets and are used as inputs to
 1457 the existing jet flavour tagger DL1r. The addition of this auxiliary training objective
 1458 removes the need for inputs from a dedicated secondary vertexing algorithm.

1459 Both of the auxiliary training objectives described here can be considered as “stepping
 1460 stones” on the way to classifying the flavour of the jet. By requiring the model to
 1461 predict the truth origin of each track and the vertex compatibility of each track-pair,
 1462 the model is guided to learn representations of the jet which are connected to the
 1463 underlying physics and therefore relevant for classifying the jet flavour.

1464 6.4.3 Architecture

1465 As discussed in the previous sections, GN1 is a graph neural network which makes
 1466 use of auxiliary training objectives in order to determine the jet flavour. A coarse
 1467 optimisation of the network architecture hyperparameters (for example number of
 1468 layers and number of neurons per layer) has been carried out in order to maximise
 1469 the flavour tagging performance, but it is likely that further dedicated optimisation
 1470 studies could lead to further performance improvements.

1471 The model architecture builds on a previous implementation of a GNN-based jet
 1472 tagger [111]. The previous approach was comprised of two separate graph neural
 1473 networks with the auxiliary tasks being performed at an intermediate stage after the

1474 first and before the second. This two stage approach was found to be unnecessary and
1475 as such GN1 simplifies the architecture into a single graph neural network with the
1476 auxiliary tasks being performed at the end, alongside the primary jet classification
1477 task. GN1 makes use of a more sophisticated graph neural network layer [119],
1478 which is described in more detail below. The changes significantly improved tagging
1479 performance and also led to a significant reduction in training time.

1480 As inputs, the model takes information about the jet and a number of associated
1481 tracks, as detailed in Section 6.4.1. The jet variables are concatenated with the
1482 variables for each track as shown in Fig. 6.3. The combined jet-track input vectors
1483 are then fed into a per-track initialisation network with three hidden layers, each
1484 containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 6.4. The
1485 track initialisation network is similar to a deep sets model [115], but does not include
1486 a reduction operation (mean or summation) over the output track representations.
1487 The initialisation network allows for initial per-track input processing without the
1488 associated parameter count cost of the graph convolutional layers described below.

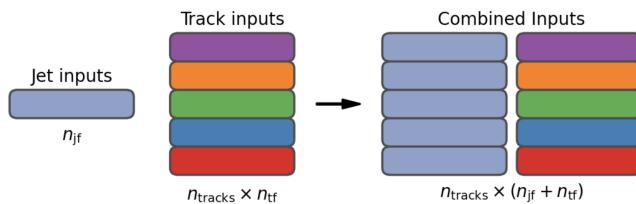


Figure 6.3: The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of n_{tracks} , where each track is described by 21 track features ($n_{\text{tf}} = 21$). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of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

1489 The outputs of the track initialisation network are used to populate the nodes of a
1490 fully connected graph, such that each node in the graph neighbours every other node.
1491 Each node h_i in the graph corresponds to a single track in the jet, and is characterised
1492 by a feature vector, also called a representation. The per-track output representations
1493 from the initialisation networks are used as the initial feature vectors of each node in
1494 the graph. In each layer of the graph network, output node representations h'_i are
1495 computed by aggregating the features of h_i and neighbouring nodes \mathcal{N}_i as described
1496 in Ref. [119]. First, the feature vectors of receiver and sender nodes are fed into two
1497 fully connected linear layers \mathbf{W}_r and \mathbf{W}_s , to produce an updated representation for

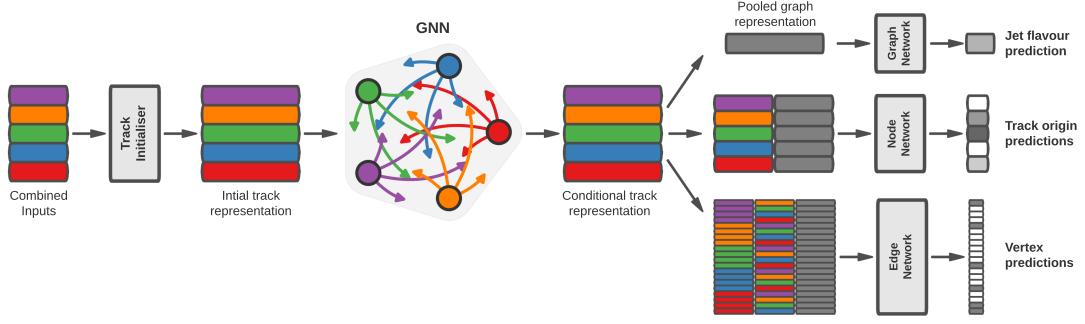


Figure 6.4: The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting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

1498 each sender and receiver node $\mathbf{W}_r h_i$ and $\mathbf{W}_s h_j$. These updated feature vectors are
 1499 used to compute edge scores $e(h_i, h_j)$ for each node pair, as in

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

1500 where, θ is a non-linear activation function, and \mathbf{a} is a learned vector. These edge
 1501 scores are then used to calculate attention weights a_{ij} for each pair of nodes using
 1502 the softmax function over the edge scores

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

1503 Finally, the updated representations for the receiver nodes h'_i are computed by taking
 1504 the weighted sum over each updated node representation $\mathbf{W}_r h_i$, with weights a_{ij}

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

1505 The set of operations described above constitute a single graph network layer. Three
 1506 such layers are stacked to construct the graph network, representing a balance
 1507 between achieving good performance and avoiding overtraining due to inflation of
 1508 the parameter count of the model. The final output from the graph neural network is
 1509 a set of per-node (i.e. per-track) feature vectors that are conditional representations
 1510 of each track given the other tracks in the jet. In order to perform the jet flavour
 1511 prediction, a flattened global representation of the jet is needed. To produce this, the
 1512 output track representations are combined using a weighted sum, where the weights
 1513 are learned during training and therefore act as a form of single-headed attention
 1514 over the different tracks. The flattened outputs from the sum are then fed into a
 1515 fully connected feedforward neural network with four layers and three outputs, one
 1516 for each jet flavour. Two other separate fully connected feedforward neural networks
 1517 are then used to independently perform the auxiliary classification objectives of GN1.
 1518 Both of the auxiliary classification tasks also take in the global representation of the
 1519 jet as inputs. A summary of the different classification networks used for the various
 1520 training objectives is shown in Table 6.2.

Table 6.2: A summary of GN1’s different classification networks used for the different training objectives. The hidden layers column contains a list specifying the number of neurons in each layer.

Network	Hidden layers	Output size
Node classification network	128, 64, 32	7
Edge classification network	128, 64, 32	1
Graph classification network	128, 64, 32, 16	3

1521 A node classification network predicts the track truth origin as defined in Table 5.1.
 1522 This network takes as inputs the features from a single output node from the graph
 1523 network and the global representation of the jet. The node network has three hidden
 1524 layers containing 128, 64 and 32 neurons respectively, and an output size of seven,
 1525 corresponding to the seven different truth origins defined in Table 5.1.
 1526 An edge classification network is used to predict whether the tracks in the track-
 1527 pair belong to a common vertex. This network takes as inputs the concatenated
 1528 representations from each pair of tracks and the global jet representation. Similar
 1529 to the node network, the edge network has three hidden layers containing 128,

1530 64 and 32 neurons respectively, and a single output, which is used to perform
 1531 binary classification of the track-pair compatibility. The output predictions for the
 1532 two auxiliary networks are used for the auxiliary training objectives discussed in
 1533 Section 6.4.2.

1534 Finally, the graph classification network is used to predict the jet flavour. This
 1535 network takes only the global jet representation as input. The graph classification
 1536 network is comprised of four fully connected hidden layers with 128, 64, 32 and 16
 1537 neurons respectively, and has three outputs corresponding to the b -, c - and light-jet
 1538 classes.

1539 6.4.4 Training

1540 The full GN1 training procedure minimises the total loss function L_{total} , defined
 1541 in Eq. (6.7). This loss is composed of three terms: L_{jet} , the categorical cross
 1542 entropy loss over the different jet flavours; L_{vertex} , the binary track-pair compatibility
 1543 cross entropy loss; and L_{track} , the categorical cross entropy loss for the track origin
 1544 prediction. L_{vertex} is computed by averaging over all intra-jet track-pairs in the batch,
 1545 and L_{track} is computed by averaging over all tracks in the batch.

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

1546 The different losses converge to different values during training, reflecting differences
 1547 in the relative difficulty of the various training objectives. The values of L_{vertex} and
 1548 L_{track} are weighted by $\alpha = 1.5$ and $\beta = 0.5$ respectively to ensure they converge to
 1549 similar values, giving them an equal weighting towards L_{total} . The values of α and β
 1550 are chosen to ensure that L_{jet} converges to a larger value than either L_{vertex} and L_{track} ,
 1551 which reflects the primary importance of the jet classification objective. It was found
 1552 that in practice the overall performance of the model was not sensitive to modest
 1553 changes in the loss weights α and β . Pre-training using L_{total} (i.e. on all tasks) and
 1554 fine tuning on only the jet classification task also did not improve performance versus
 1555 the described standard setup, indicating that the auxiliary tasks are not in direct
 1556 competition with the jet classification task. As there was a large variation in the

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

GN1 can be trained with either the node or edge networks (and their corresponding auxiliary tasks), or both, removed, as discussed in Section 6.5.4. In such cases, the corresponding losses L_{vertex} and L_{track} are also removed from the calculation of the overall loss L_{total} . The performance of the resulting models provides a useful indication of the benefit of including the auxiliary tasks to the primary jet classification objective.

GN1 is trained for 100 epochs on 4 NVIDIA V100 GPUs. Each epoch takes approximately 25 mins to complete over the training sample of 30 million jets described in Section 6.3.1. The Adam optimiser [120] with an initial learning rate of 1e–3, and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically the validation loss, calculated on 500k jets, became stable after around 60 epochs. The epoch that minimized the validation loss was used for evaluation. GN1 has been integrated into the ATLAS software [52] using ONNX [121]. The test sample jet flavour predictions scores are computed using the ATLAS software stack as a verification of this process.

6.5 Results

The GN1 tagger is evaluated both as a b -tagging and c -tagging algorithm in Section 6.5.1 and Section 6.5.2 respectively. Evaluation is performed separately on both jets in the $t\bar{t}$ sample with $20 < p_{\text{T}} < 250 \text{ GeV}$ and jets in the Z' sample with $250 < p_{\text{T}} < 5000 \text{ GeV}$. The performance of the model is compared to the DL1r tagger [108, 109], which has been retrained on 75 million jets from the same samples as GN1. The input RNNIP tagger [113] to DL1r has not been retrained. As discussed, each tagger predicts the probability that a jet belongs to the b -, c - and light-classes. To use the model for b -tagging, these probabilities are combined into a single score D_b , which is defined as

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

1586 where f_c is a free parameter that determines the relative weight of p_c to p_l in the
 1587 score D_b , controlling the trade-off between c - and light-jet rejection performance.
 1588 The choice of f_c is arbitrary, with the different optimised values reflecting the relative
 1589 c - versus light-jet rejection performance of the various taggers. This parameter is
 1590 set to a value of $f_c = 0.018$ for the DL1r model, obtained through an optimisation
 1591 procedure designed to maximise the c - and light-jet rejection of DL1r described in
 1592 Ref. [108]. Based on a similar optimisation procedure, a value of $f_c = 0.05$ is used
 1593 for the GN1 models. A fixed-cut working point (WP) defines the corresponding
 1594 selection applied to the tagging discriminant D_b in order to achieve a given efficiency
 1595 on the inclusive $t\bar{t}$ sample.

1596 The technical implementation of GN1 in the ATLAS software stack requires any jet
 1597 with zero or one associated tracks to be classified as a light-jet. The impact of this
 1598 on the tagging performance of GN1 was found to be negligible, with 0.12% of b -jets
 1599 in the $t\bar{t}$ sample and 0.02% of b -jets in the Z' sample affected. Of those, 89% of the
 1600 b -jets in the $t\bar{t}$ sample and 98% of the b -jets in the Z' sample are also classified as
 1601 light-jets by DL1r at the 70% $t\bar{t}$ WP.

1602 A comparison of the b -tagging discriminant D_b between DL1r and GN1 is shown in
 1603 Fig. 6.5. The shapes of the D_b distributions are generally similar for b -, c - and light-
 1604 jets between both models, however, GN1 shifts the b -jet distribution to higher values
 1605 of D_b in the regions with the greatest discrimination. The GN1 c -jet distribution is
 1606 also shifted to lower values of D_b when compared with DL1r, enhancing the separation
 1607 and indicating that GN1 is improving c -jet rejection when compared with DL1r.

1608 6.5.1 b -tagging Performance

1609 The performance of b -tagging algorithms is quantified by their ability to reject c -
 1610 and light-jets for a given b -jet selection efficiency, or working point (WP). In order
 1611 to compare the b -tagging performance of the different taggers for the b -jet tagging
 1612 efficiencies in the range typically used by analyses, the corresponding c - and light-jet

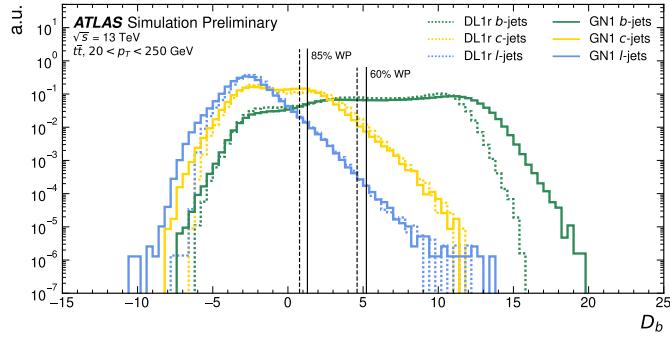


Figure 6.5: Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs used by analyses. A value of $f_c = 0.018$ is used in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

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

For jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV, GN1 demonstrates considerably better c - and light-jet rejection when compared with DL1r across the full range of b -jet tagging efficiencies studied. The relative improvement is strongly dependent on the b -jet tagging efficiency under study. The largest improvements are found at lower b -jet tagging efficiencies. At a b -jet tagging efficiency of 70%, the c -rejection improves by a factor of ~ 2.1 while the light-jet rejection improves by a factor of ~ 1.8 with respect to DL1r. For high- p_T jets in the Z' sample with $250 < p_T < 5000$ GeV, GN1 also brings a significant performance improvement with respect to DL1r across the

range of b -jet tagging efficiencies studied. Again, the largest relative improvement in performance comes at the lower b -jet tagging efficiencies. At a b -jet efficiency of 30%, GN1 improves the c -rejection with respect to DL1r by a factor of ~ 2.8 and the light-jet rejection by a factor of ~ 6 . The performance comparison at lower b -jet tagging efficiencies is made more difficult due to the increased statistical uncertainties which result from the high rejection of background. It is estimated that for a b -jet tagging efficiency of 70% in the $t\bar{t}$ sample, $\sim 5\%$ ($\sim 30\%$) of the relative improvement in the c -jet (light-jet) rejection comes from loosening the track selection and for a b -jet tagging efficiency of 30% in the Z' the corresponding number is $\sim 10\%$ for both c -jets and light-jets. Given that GN1 exploits the low-level detector information in a more complete and sophisticated way than DL1r, further studies are needed to confirm that the performance gain observed in these simulated samples is also observed in experimental data.

The GN1Lep variant of GN1 demonstrates further improved performance with respect to the baseline model. This demonstrates the additional jet flavour discrimination power provided by the leptonID track input. For jets in the $t\bar{t}$ sample, the relative c -rejection improvement with respect to DL1r at the 70% b -jet WP increases from a factor of ~ 2.1 for GN1 to a factor of ~ 2.8 for GN1Lep. The improvement in light-jet rejection also increases from a factor of ~ 1.8 to ~ 2.5 at the same WP. For jets in the Z' sample, the relative c -rejection (light-jet rejection) improvement with respect to DL1r increases from a factor of ~ 2.8 to ~ 3 (~ 6 to ~ 7.5) at a b -jet tagging efficiency of 30%. As shown in Fig. 6.8, the greatest improvement of GN1Lep over GN1 is seen at lower transverse momenta.

In general, the performance of all the taggers is strongly dependent on the jet p_T . This is due to the increased multiplicity and collimation of tracks, and the displaced decays that result from within the heavy flavour jets. These factors are described in more detail in Chapter 4. Together, they contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a general degradation in quality of tracks inside the core of a jet, which in turn reduces the jet tagging performance.

In order to study how the tagging performance changes as a function of the jet p_T , the b -jet tagging efficiency as a function of p_T for a fixed light-jet rejection of 100 in each bin is shown in Fig. 6.8. For jets in the $t\bar{t}$ sample, at a fixed light-jet rejection of 100, GN1 improves the b -jet tagging efficiency by approximately 4% across all the jet p_T bins. Meanwhile, GN1Lep again demonstrates improved performance with

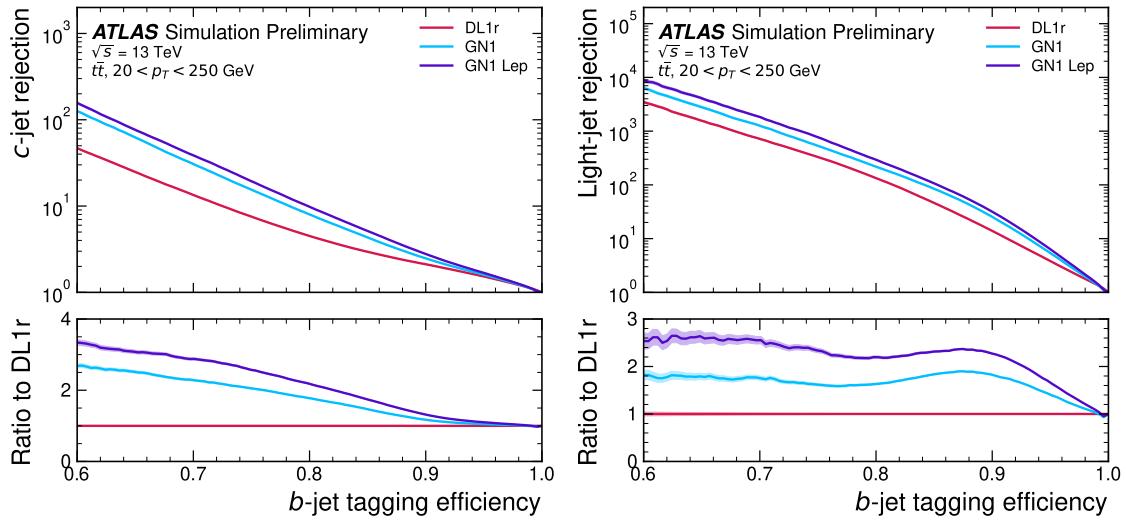


Figure 6.6: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 75\%$, the light-jet rejection becomes so large that the effect of the low number of jets is visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions of phase space.

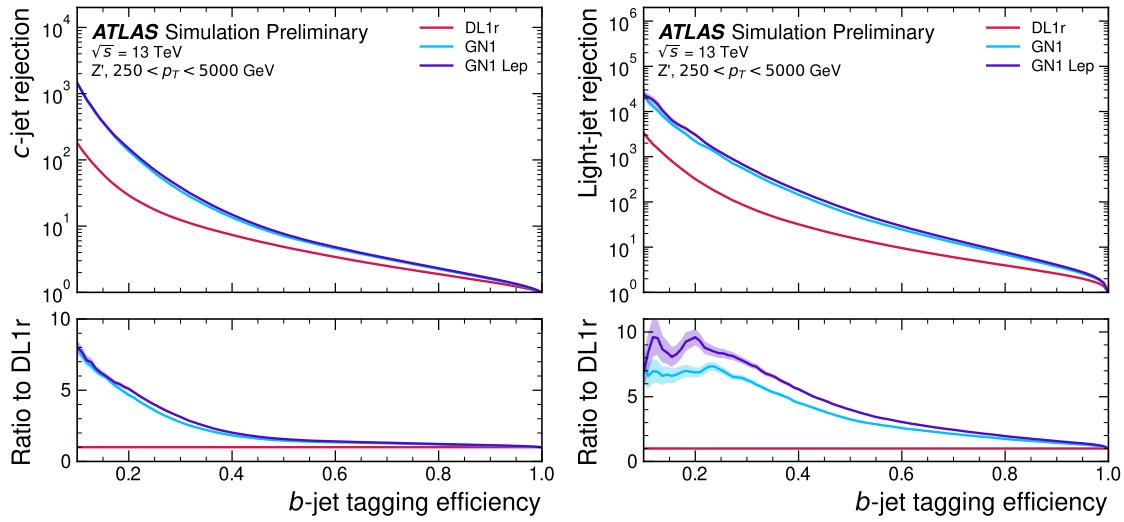


Figure 6.7: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used 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. At b -jet tagging efficiencies less than $\sim 20\%$, the light-jet rejection becomes so large that the effect of the low number of jets is visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions of phase space.

respect to GN1, in particular at lower p_T . The relative increase in the b -jet tagging efficiency increases from 4% to 8% with respect to DL1r. For jets in the Z' sample, GN1 again outperforms DL1r across the entire jet p_T range studied. The largest relative improvement in performance is found at the highest transverse momenta of jet $p_T > 2$ TeV, and corresponds to an approximate factor of 2 improvement in efficiency with respect to DL1r.

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

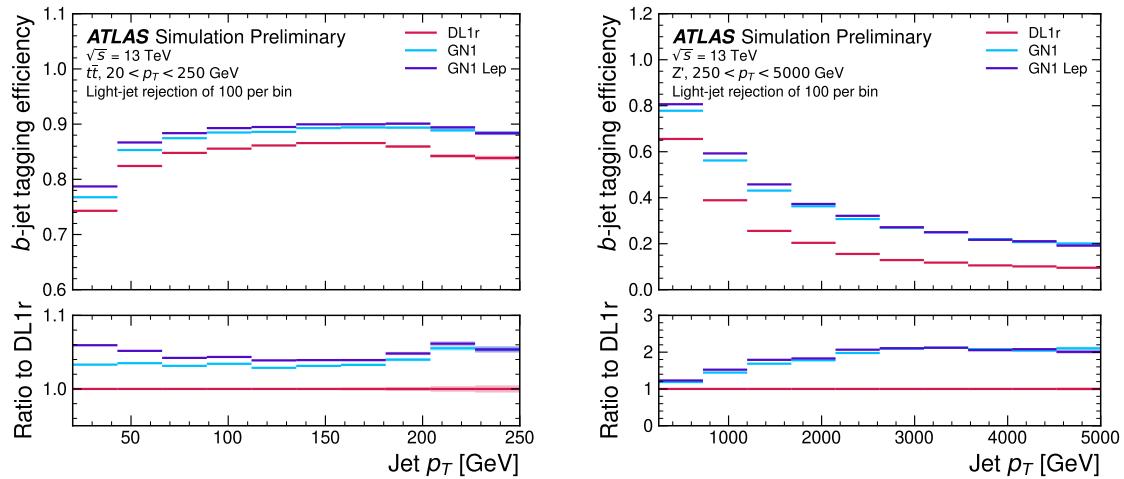


Figure 6.8: The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. GN1 demonstrates improved performance with respect to DL1r accross 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

1684 approach used for b -tagging. The model output probabilities are combined into a
1685 single score D_c , which is defined similarly to Eq. (6.8) as

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

1686 A value of $f_b = 0.2$ is used for all models, based on the same optimisation procedure
1687 that was used for the b -tagging use case. Similar to Section 6.5.1, the different
1688 taggers are compared to one another by scanning through a range of c -jet tagging
1689 efficiencies and plotting the corresponding b - and light-jet rejection rates. As in
1690 Section 6.5.1, the WPs are defined using jets in the $t\bar{t}$ sample. Standard c -jet tagging
1691 efficiency WPs used by physics analyses are significantly lower than the b -tagging
1692 WPs in order to maintain reasonable b - and light-jet rejection rates. This is reflected
1693 in the range of c -jet tagging efficiencies used in c -tagging plots such as Figs. 6.9
1694 and 6.10. Fig. 6.9 displays the c -tagging performance of the models on the jets in
1695 the $t\bar{t}$ sample. GN1 is shown to perform significantly better than DL1r. Similar to
1696 the b -tagging case, the b - and light-jet rejection improve most at lower c -jet tagging
1697 efficiencies, with both background rejections increasing by a factor of 2 with respect
1698 to DL1r at a c -jet tagging efficiency of 25%. GN1Lep again outperforms GN1, though
1699 the improvements with respect to the baseline model are more modest, with the
1700 b -rejection (light-jet rejection) relative improvement increasing from a factor of 2 to
1701 2.1 (2 to 2.3) at the 25% c -jet WP. Fig. 6.10 shows the c -tagging performance on the
1702 jets in the Z' sample with $250 < p_T < 5000$ GeV. Both GN1 and GN1Lep perform
1703 similarly, improving the b -rejection by 60% and the light-jet rejection by a factor of
1704 2 at the 25% c -jet WP.

1705 6.5.3 Jet Display Diagrams

1706 The auxiliary training objectives of GN1 allow for improved model interpretability,
1707 which is especially important for a monolithic approach as the low level taggers,
1708 which provide useful physical insight, are no longer present. Figs. 6.11 and 6.12
1709 show a comparison of the true origin and vertexing information compared with
1710 the predicted values from GN1, SV1 and JetFitter. Such comparisons can be used

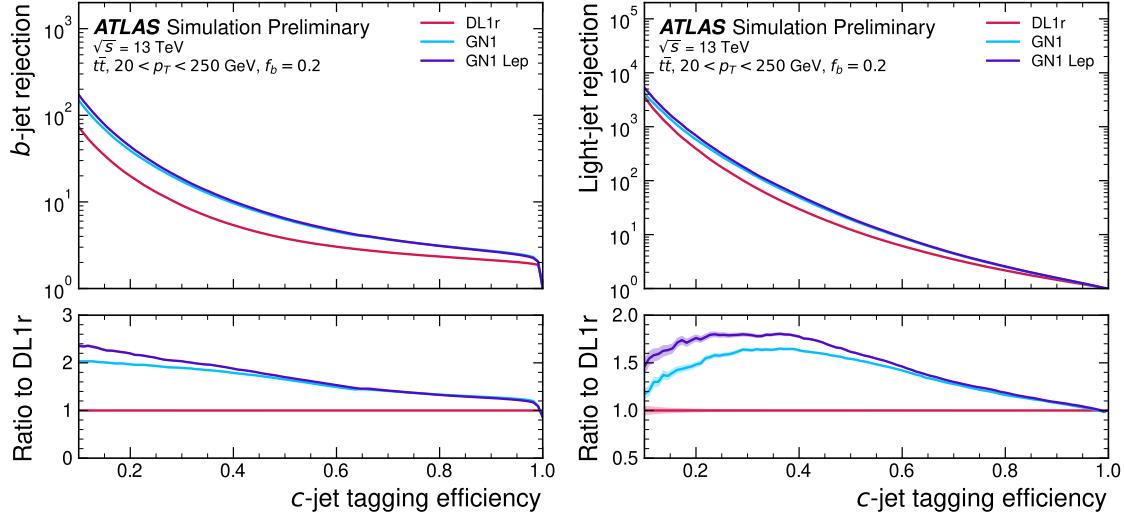


Figure 6.9: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. At c -jet tagging efficiencies than $\sim 25\%$, the light-jet rejection becomes so large that the effect of the low number of jets is visible. The lower x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.

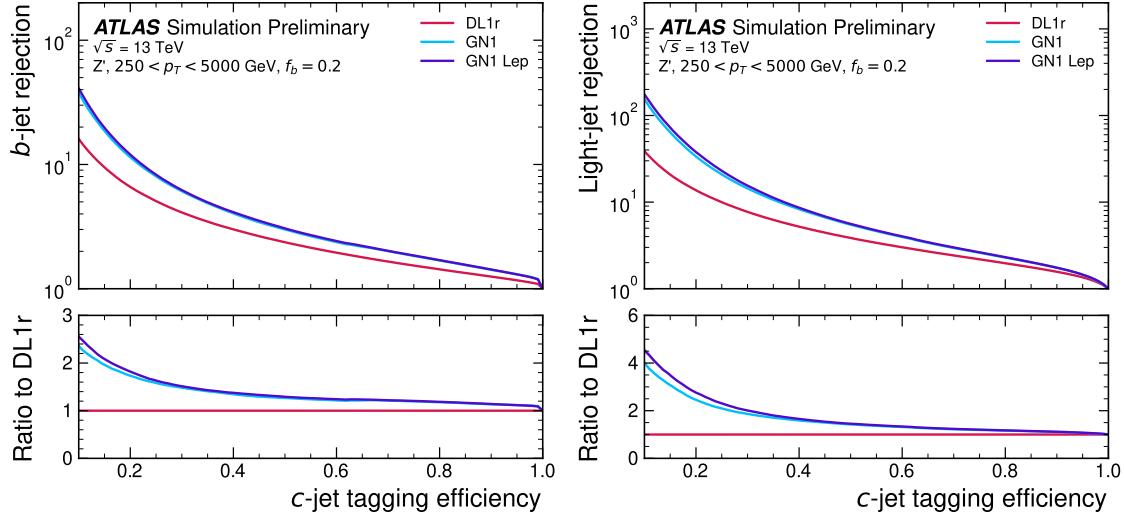


Figure 6.10: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000 \text{ GeV}$. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The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

1711 to provide an indication that GN1 recovers the intended representation of the jet
1712 structure, and may also help to identify limitations of the model.

1713 The diagrams show the truth (left) and predicted (right) structure of a b -jet. The
1714 shaded black boxes show the groupings of tracks into different vertices. In Fig. 6.11,
1715 GN1 correctly groups the three primary tracks as having come from the primary
1716 vertex. The b -hadron and $b \rightarrow c$ -hadron decay vertices are also correctly predicted,
1717 and the origin of the tracks in each is correct. There is a single OtherSecondary track
1718 which GN1 incorrectly predicts as having come pileup. Meanwhile SV1 (by design)
1719 merges the two heavy flavour decay vertices, but incorrectly includes a track from
1720 the primary vertex. JetFitter reconstructs two vertices, one which is a combination
1721 of two tracks from the different truth vertices and two other single track vertices in
1722 each of the heavy flavour vertices. GN1 also predicts the flavour of the jet with a
1723 high degree of certainty.

1724 Similarly, Fig. 6.11 shows that, at least in some cases, GN1 is able to accurately
1725 predict the origin and vertex information of tracks inside a jet. The pileup tracks
1726 and primary vertex tracks are correctly identified, and the heavy flavour decay tracks
1727 are also correctly identified with the exception of one of the b -hadron decay tracks.
1728 Again, SV1 is overly aggressive by merging the two heavy flavour decay vertices
1729 along with a track from pileup, while JetFitter shows signs of being underconstrained
1730 by reconstructing two single track vertices, one with a pileup track and one with a
1731 track from a $b \rightarrow c$ -hadron decay.

1732 6.5.4 Ablations

1733 Ablation studies (the removal of certain components of a given model in order to
1734 study the impact of that component) are carried out to determine the
1735 importance of the auxiliary training objectives of GN1 to the overall performance.
1736 The “GN1 No Aux” variant retains the primary jet classification objective, but
1737 removes both track classification and vertexing auxiliary objectives (see Section 6.4.2)
1738 and correspondingly only minimises the jet classification loss. The “GN1 TC” variant
1739 includes track classification objective but not the vertexing objective. Finally, the
1740 “GN1 Vert” includes the vertexing objective, but not the track classification objective.

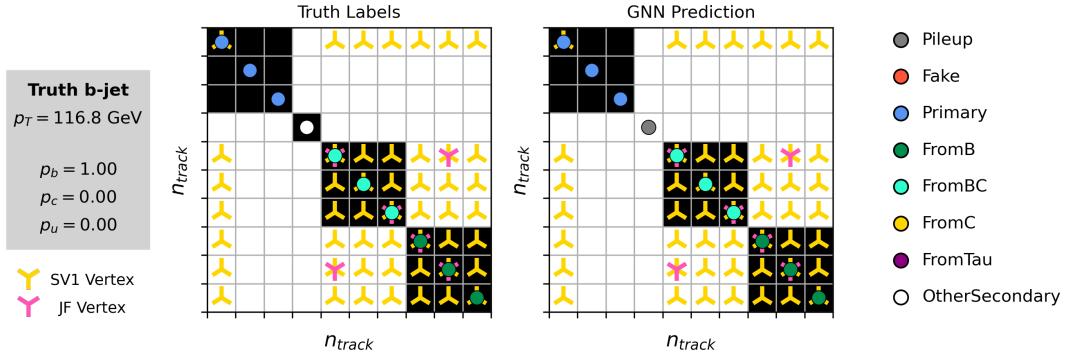


Figure 6.11: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities p_b , p_c and p_u are rounded to three significant figures. GN1 improves over SV1 and 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.

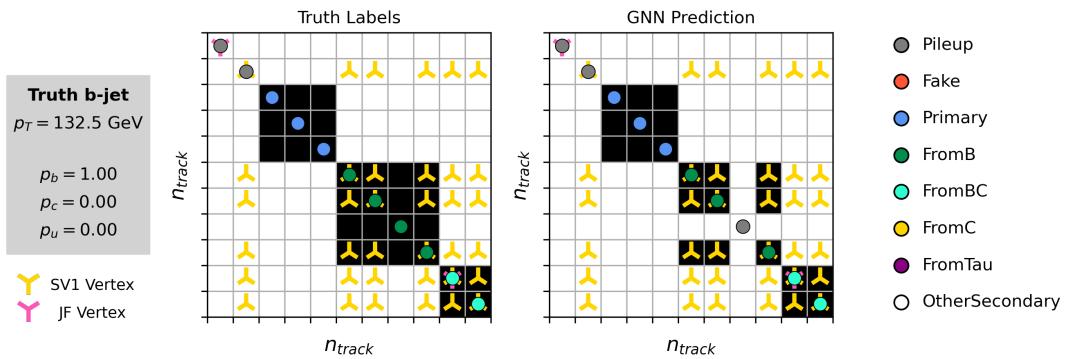


Figure 6.12: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities p_b , p_c and p_u are rounded to three significant figures.

1741 For jets in both the $t\bar{t}$ and Z' samples, a general trend is observed that the models
1742 trained without one or both of the auxiliary objectives results in significantly reduced
1743 c - and light-jet rejection when compared with the baseline GN1 model. This result
1744 is shown clearly in Figs. 6.13 and 6.14. For jets in the $t\bar{t}$ sample, the performance
1745 of GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert perform similarly
1746 to each other with improvements in the c -rejection of 80% and improvements in
1747 the light-jet rejection of 75% at a b -jet tagging efficiency of 60%. For jets in the Z'
1748 sample meanwhile, the GN1 No Aux model already shows a clear improvement in c -
1749 and light-jet rejection when compared with DL1r at lower b -jet tagging efficiencies.
1750 Similar to jets in the $t\bar{t}$ sample, GN1 TC and GN1 Vert perform similarly, and bring
1751 large gains in background rejection when compared with GN1 No Aux, but the
1752 combination of both auxiliary objectives yields the best performance.

1753 It is notable that the GN1 No Aux model matches or exceeds the performance of
1754 DL1r without the need for inputs from the low-level algorithms. This indicates that
1755 the performance improvements enabled by the improved neural network architecture
1756 used in GN1 appear to be able to compensate for the removal of the low-level
1757 algorithm inputs. The GN1 TC and GN1 Vert variants each similarly outperform
1758 DL1r, demonstrating that both contribute to the overall high performance of the
1759 baseline model. The overall best performing model is the full version of GN1 trained
1760 with both auxiliary objective, demonstrating that the both auxiliary objectives are
1761 complementary.

1762 6.5.5 Inclusion of Low-Level Vertexing Algorithms

1763 As already mentioned, GN1 does not include any inputs from the low-level tagging
1764 algorithms, including the vertexing algorithms SV1 and JetFitter [60]. Since these
1765 algorithms are known to play a key role in contributing to the performance of DL1r, it
1766 was necessary to study whether their inclusion in GN1 resulted in further performance
1767 improvements. In a dedicated training of GN1 the SV1 and JetFitter tagger outputs
1768 were added to the GN1 jet classification network as an input, similar to how they
1769 are used in DL1r. These outputs include information on the reconstructed vertices,
1770 including the number of vertices, and the mass, displacement, and other properties of
1771 the highest ranking reconstructed vertex. In addition, the index of the reconstructed
1772 SV1 or JetFitter vertices were included as two track-level inputs to GN1. These

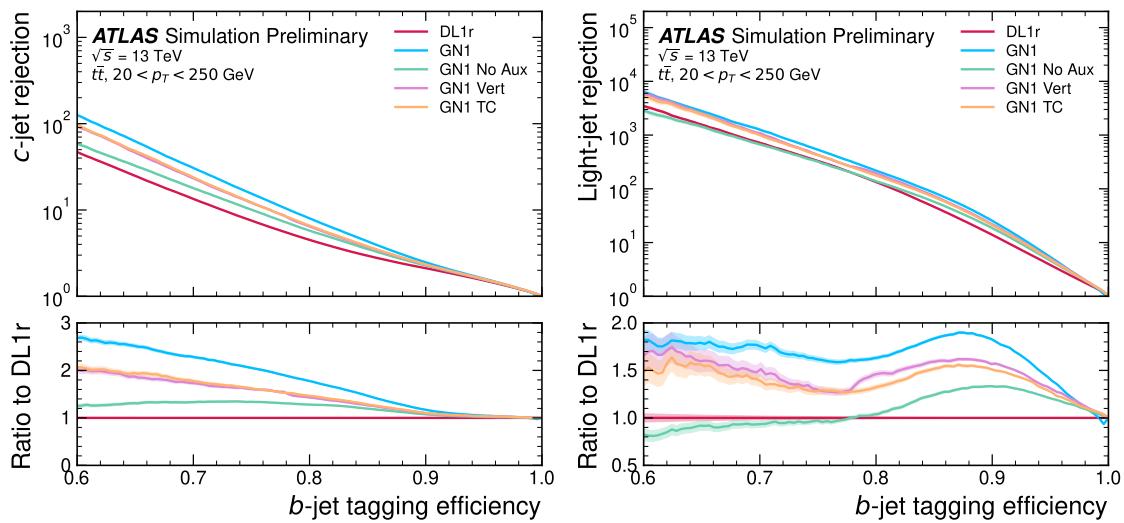


Figure 6.13: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 65\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.

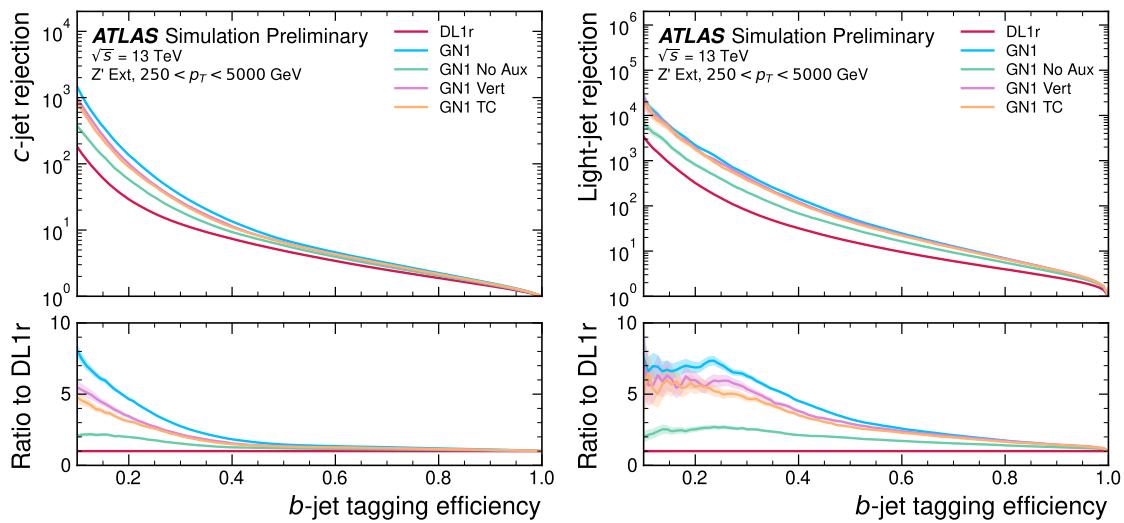


Figure 6.14: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 25\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.

1773 indices were also used to construct an edge feature for the edge classification network,
1774 which was given a value of one if the track-pair were from a common reconstructed
1775 V1 or JetFitter vertex, and zero otherwise. The jet classification performance of this
1776 GN1 model was not significantly different to the baseline model, and in some cases
1777 the performance was slightly reduced. As GN1 does not benefit from the inclusion
1778 of information from SV1 and JetFitter it can function as a highly performant
1779 standalone tagger that does not require (beyond retraining) any manual optimisation
1780 to achieve good performance in a wide range of phase spaces. A dedicated look at
1781 the vertexing performance of GN1 with some comparisons to SV1 and JetFitter is
1782 found in Section 6.5.6

1783 6.5.6 Vertexing Performance

1784 From the track-pair vertex prediction described in Section 6.4.2, tracks can be
1785 partitioned into compatible groups representing vertices (see [111]). As such, GN1
1786 is able to be used to perform vertex “finding”, but not vertex “fitting”, i.e. the
1787 reconstruction of a vertex’s properties, which currently still requires the use of a
1788 dedicated vertex fitter. In order to study the performance of the different vertexing
1789 tools inside b -jets, the truth vertex label of the tracks, discussed in Section 6.4.2, are
1790 used. To estimate the efficiency with which GN1 manages to find vertices inclusively,
1791 vertices from GN1 containing tracks identified as coming from a b -hadron are merged
1792 together and compared to the inclusive truth decay vertices that result from a
1793 b -hadron decay (where if there are multiple distinct truth vertices from a b -hadron
1794 decay they are also merged together). Vertices are compared with the target truth
1795 vertex and the number of correctly and incorrectly assigned tracks is computed.
1796 Since secondary vertex information is only recovered for reconstructed tracks, an
1797 efficiency of 100% here denotes that all possible secondary vertices are recovered
1798 given the limited track reconstruction efficiency. A vertex is considered matched if
1799 it contains at least 65% of the tracks in the corresponding truth vertex, and has a
1800 purity of at least 50%. GN1 manages to achieve an inclusive reconstruction efficiency
1801 in b -jets of $\sim 80\%$, demonstrating that it effectively manages to identify the displaced
1802 vertices from b -hadron decays.

1803 In order to study the performance of the different vertexing tools inside b -jets, the
1804 truth vertex label of the tracks, discussed in Section 6.4.2, is used. The reconstructed

1805 vertices from GN1, SV1 and JetFitter are compared to the target truth vertices in
1806 order to calculate the efficiencies of the different vertexing tools. Since secondary
1807 vertex information is only recovered for reconstructed tracks, an efficiency of 100%
1808 here denotes that all possible secondary vertices are recovered given the limited track
1809 reconstruction efficiency.

1810 There are several caveats to a comparison of the vertexing tools which are a result
1811 of the different approaches they take to vertexing. SV1 and JetFitter are designed
1812 to only find secondary vertices in the jet, whereas GN1 is also trained to determine
1813 which tracks in the jet belong to the primary vertex (the vertex of the hard scatter
1814 pp interaction). To account for this the GN1 vertex with the largest number of
1815 predicted primary tracks is excluded from the vertex finding efficiency calculation.
1816 While JetFitter and GN1 aim to resolve each displaced vertex inside the jet, such
1817 that secondary vertices from b -hadron decays are found separately to tertiary vertices
1818 from $b \rightarrow c$ decay chains, SV1 by design attempts to find a single inclusive vertex
1819 per jet. This inclusive vertex groups inclusive b -hadron decays. These are tracks
1820 from the b -hadron decay itself (FromB) and tracks from $b \rightarrow c$ decays (FromBC).
1821 In order to fairly compare the performance if the different tools, both the exclusive
1822 and inclusive vertex finding efficiency is studied. For the exclusive vertex finding
1823 case JetFitter and GN1 can be directly compared, while a comparison with SV1 is
1824 not possible due to aforementioned design constraints. The inclusive vertex finding
1825 performance of all three tools can be compared using the procedure outlined below.

1826 The starting point for the secondary vertex finding efficiency in both the exclusive
1827 and inclusive cases is to select truth secondary vertices are those containing only
1828 inclusive b -hadron decays to be considered as initial targets. For exclusive vertex
1829 finding, these truth secondary vertices can be used directly as the denominator for the
1830 efficiency calculation. Meanwhile for the inclusive efficiency all such truth secondary
1831 vertices in the jet are merged into a single inclusive target vertex. Correspondingly,
1832 for the inclusive vertex finding case, the vertices found by JetFitter are merged into
1833 a single vertex, and the vertices found by GN1 with at least one predicted inclusive
1834 b -hadron decay track are also merged similarly. SV1 does not require any vertex
1835 merging.

1836 Next, in both cases for each truth secondary vertex, vertices in the jet found by the
1837 different vertexing tools are compared with the target truth vertex. The number
1838 of correctly and incorrectly assigned tracks is computed. In order to call a vertex

efficient, it is required to contain at least 65% of the tracks in the corresponding truth vertex, and to have a purity of at least 50%. Single track vertices are required to have a purity of 100%. Additionally, for GN1 only, at least one track in the vertex is required to have a predicted heavy flavour origin.

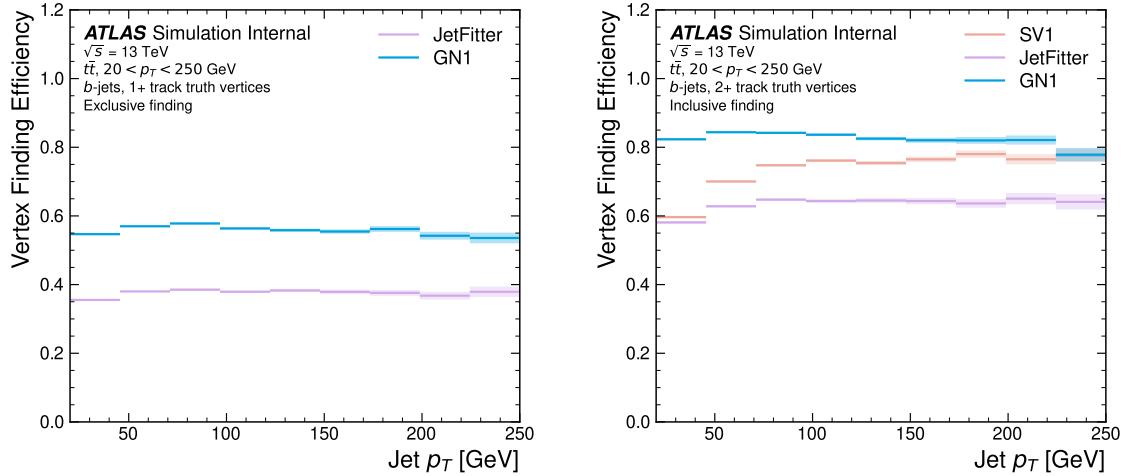


Figure 6.15: Vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50%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

Vertex finding efficiencies for b -jets in the $t\bar{t}$ sample are displayed as a function of p_T separately for the inclusive and exclusive approaches in Fig. 6.15. For b -jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$, the exclusive vertex finding efficiency of JetFitter and GN1 is relatively flat as a function of p_T . Of the truth secondary vertices in this p_T region, JetFitter efficiently finds approximately 40% and GN1 finds approximately 55%. When finding vertices inclusively the vertex finding efficiency is generally higher. An increased dependence on p_T is also visible for JetFitter and SV1. As the jet p_T increases from 20 GeV to 100 GeV, the efficiency of JetFitter increases from 55% to 65%. In the same range, the efficiency of SV1 increases from 55% to 75%. GN1 displays less dependence on p_T than JetFitter and SV1, efficiently finding upwards of 80% of vertices in b -jets in this p_T region. For b -jets with $p_T > 100 \text{ GeV}$, JetFitter finds approximately 65% of vertices, SV1 finds approximately 75% of vertices, and GN1 finds approximately 80% of 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 30% efficiency. Fig. 6.16 compares the exclusive vertex finding efficiencies of JetFitter and GN1 for multitrack vertices. 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.

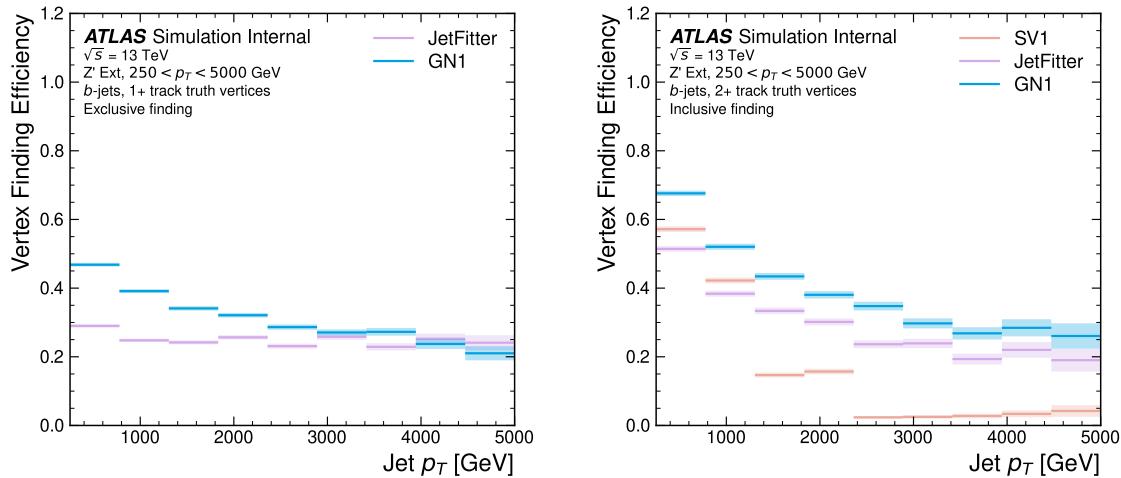


Figure 6.16: Inclusive vertex finding efficiency for multitrack truth vertices in b -jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T . Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly.

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 based on a standard multi-class feed-forward classification network is trained on the same tracks used for the baseline GN1 training. The model uses precisely the same concatenated track-and-jet inputs as used by GN1 (see Section 6.4.1),

1873 but processes only a single track at a time, meaning it cannot take into account
1874 the correlations between tracks when determining the track origin. The model is
1875 made up of five densely connected linear layers with 200 neurons in each layer. The
1876 performance of the model was found to be unsensitive to changes in the network
1877 structure.

1878 To measure the track classification performance, the area under the curve (AUC)
1879 of the receiver operating characteristic (ROC) curve is computed for each origin
1880 class, using a one-versus-all classification approach. The AUCs for the different
1881 truth origins are averaged using both an unweighted and a weighted mean. The
1882 unweighted mean treats the performance of each class equally, while the weighted
1883 mean uses as a weight the relative abundance of tracks of each class. Table 6.3
1884 demonstrates clearly that GN1 outperforms the MLP both at $20 < p_T < 250 \text{ GeV}$
1885 for jets in the $t\bar{t}$ sample and at $250 < p_T < 5000 \text{ GeV}$ for jets in the Z' sample. For
1886 example, GN1 can reject 65% of fake tracks in jets in the $t\bar{t}$ sample, while retaining
1887 more than 99% of good tracks (i.e. those tracks which are not fake). The GN1 model
1888 has two advantages over the MLP which can explain the performance improvement.
1889 Firstly, the graph neural network architecture enables the sharing of information
1890 between tracks as discussed in Section 6.4.3. This is likely to be beneficial since the
1891 origins of different tracks within a jet are correlated. Secondly, the jet classification
1892 and vertexing objectives may be complementary to the track classification objective,
1893 and so the track classification performance is improved by the combined training of
1894 complementary objectives.

1895 Fig. 6.17 shows the track origin classification ROC curves for the different track
1896 origins for jets in both the $t\bar{t}$ and Z' samples. In order to improve visual readability
1897 of the plot, the curves for the heavy flavour truth origins (FromB, FromBC and
1898 FromC) have been combined (weighted by their relative abundance), as have the
1899 Primary and OtherSecondary origins. In jets in both the $t\bar{t}$ and Z' samples, the AUC
1900 of all the different origin groups exceeds 0.9, representing strong overall classification
1901 performance. In both samples fake tracks are the easiest to classify, followed by
1902 pileup tracks. The FromC tracks which are c -hadron decay products, are the hardest
1903 to classify, possibly due to their similarity to both fragmentation tracks and b -hadron
1904 decay tracks, depending on the c -hadron species in question.

		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.

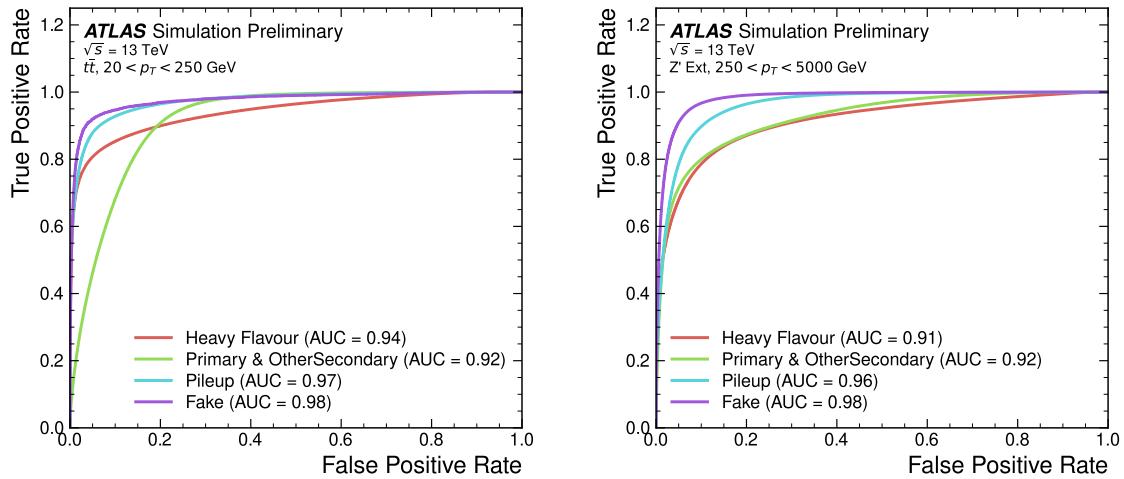


Figure 6.17: ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right). The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the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

6.5.8 Looser Track Selection

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

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

where $N_{\text{shared}}^{\text{Pix}}$ is the number of shared pixel hits and $N_{\text{shared}}^{\text{SCT}}$ is the number of shared SCT modules on a track. The nominal cut used elsewhere in this thesis is $N_{\text{shared}}^{\text{Si}} < 2$. As the rate of shared hits is significantly higher for b -hadron decay tracks than for other tracks, this cut rejects a significant proportion of these tracks. Figs. 6.18 and 6.19 show the result of training the GN1 tagger with the full relaxation of this cut, i.e. allowing tracks with any number of shared hits. The shared hit requirements applied by the ambiguity solver as part of track reconstruction (see Section 3.4.1) are still applied. In addition, the maximum allowed value of d_0 is increased from 3.5 mm to 5.5 mm.

6.6 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 $20 < p_{\text{T}} < 250$ 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 $250 < p_{\text{T}} < 5000$ GeV, GN1 improves the c -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

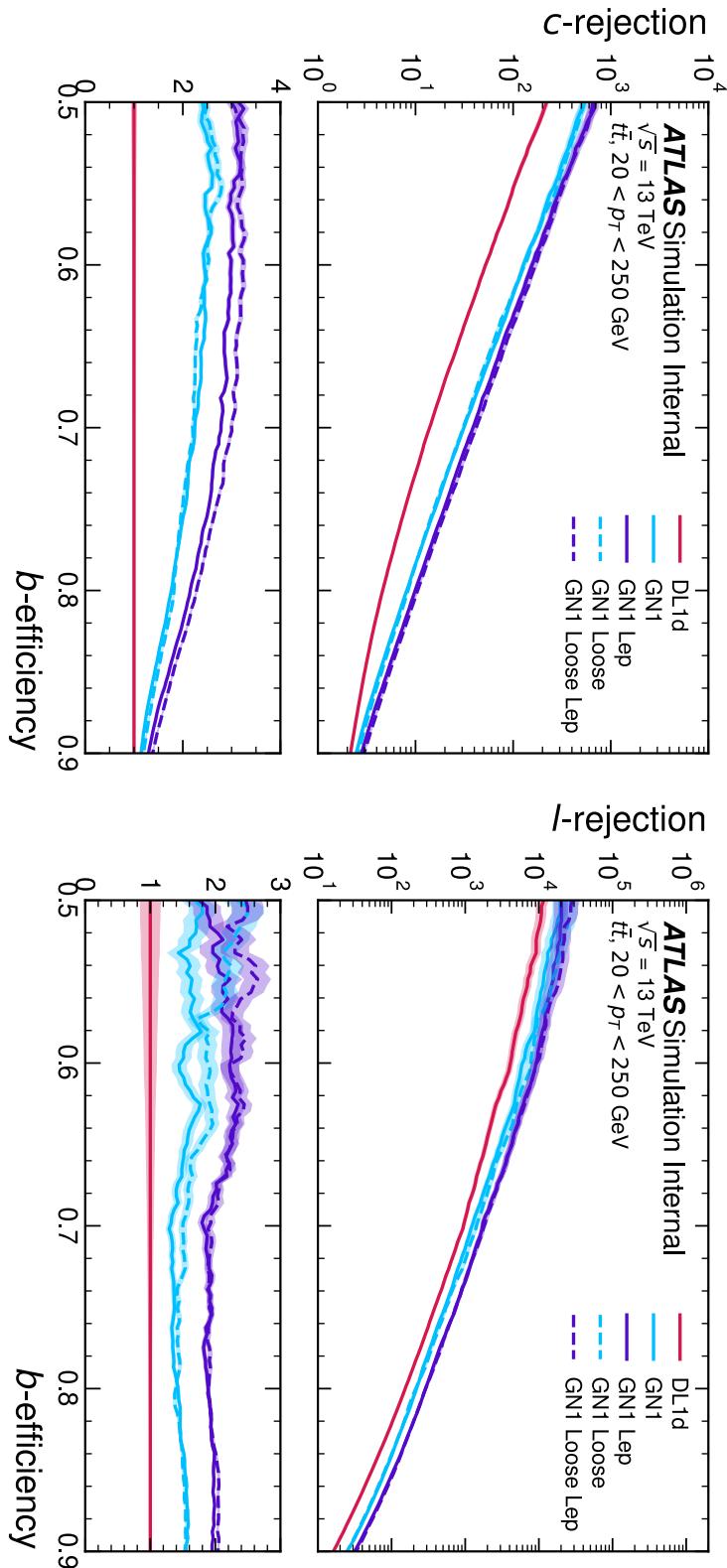


Figure 6.18: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. A value 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.

1933 time-consuming retuning of the low-level tagging algorithms. The model is also
1934 simpler to maintain and study due to the reduction of constituent components.

1935 GN1 demonstrates improved track classification performance when compared with
1936 a simple per-track MLP and an efficiency of $\sim 80\%$ for inclusive vertex finding in
1937 b -jets. The model is also able to perform vertex finding, and preliminary studies
1938 suggest it outperforms previous manually optimised approaches. The auxiliary track
1939 classification and vertex finding objectives are shown to significantly contribute to
1940 the performance in the jet classification objective, and, along with the more advanced
1941 graph neural network architecture, are directly responsible for the improvement over
1942 DL1r.

1943 Further improvements in the b - and c -tagging performance are likely possible with
1944 a more thorough optimisation of the model architecture, and the integration of
1945 additional information from other parts of the ATLAS detector. The addition of
1946 other auxiliary training objectives, such as the truth b -hadron decay radius and
1947 transverse momentum, may also yield additional performance gains.

1948 Additional future work includes the verification of the performance of GN1 on collision
1949 data, and the full calibration of the model so it can be used by analyses. The flexible
1950 nature of the model means it can also be readily applied to other related problems
1951 outside of standard b - and c -tagging applications, for example $X \rightarrow bb$ and $X \rightarrow cc$
1952 tagging. The model could also be repurposed as a pileup jet tagger, or general
1953 primary and secondary vertexing tool.

1954 The model also demonstrates strong performance for the HLLHC ??.

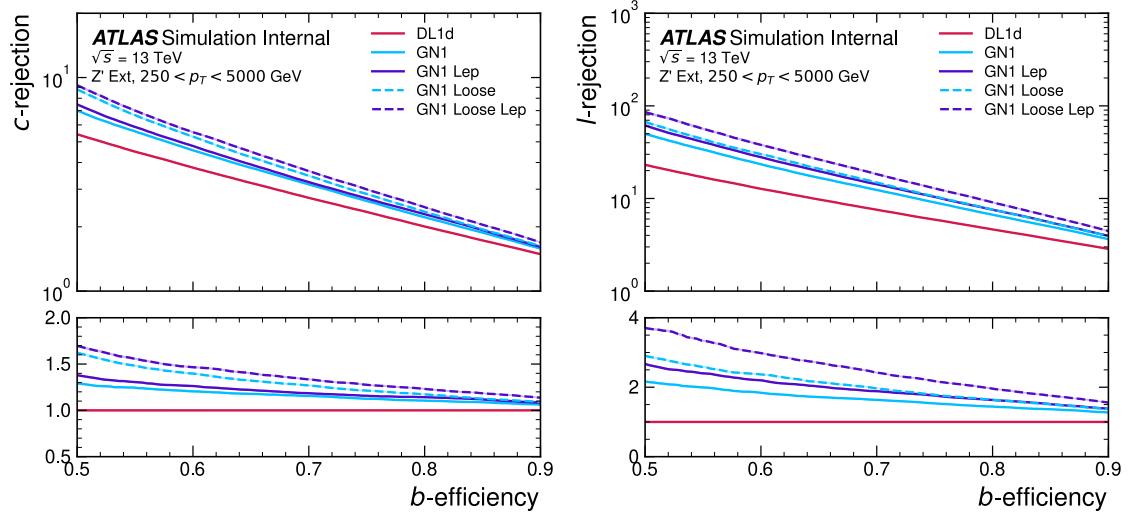


Figure 6.19: The *c*-jet (left) and light-jet (right) rejections as a function of the *b*-jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. A value of $f_c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

1955 - paper aux mat - paper internal note - bulk it out

₁₉₅₆ **Chapter 7**

₁₉₅₇ **Boosted VHbb Analysis**

₁₉₅₈ The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [12, 13],
₁₉₅₉ is predicted by the standard model to decay primarily to a pair of b -quarks, with a
₁₉₆₀ branching factor of 0.582 ± 0.007 for $m_H = 125\text{ GeV}$ [24]. Observation of this decay
₁₉₆₁ mode was reported by ATLAS [105] and CMS [26] in 2018, establishing the first
₁₉₆₂ direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks
₁₉₆₃ (see Section 2.2.2). The $H \rightarrow b\bar{b}$ process is also important for constraining the total
₁₉₆₄ decay width of the Higgs [122].

₁₉₆₅ Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion
₁₉₆₆ as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet
₁₉₆₇ background and so overall sensitivity to the Higgs is low. The QCD multijet
₁₉₆₈ background refers to events containing one or more strongly produced jets which are
₁₉₆₉ not the decay product of heavy resonances, for example $g \rightarrow q\bar{q}$. The gluon-gluon
₁₉₇₀ fusion channel contains to leading order only jets in the final state, and therefore
₁₉₇₁ it is extremely difficult to distinguish signal events from the overwhelming multijet
₁₉₇₂ background. The $H \rightarrow b\bar{b}$ observation therefore searched for Higgs bosons produced
₁₉₇₃ in association with a vector boson V (where V can be a W or Z boson) which
₁₉₇₄ subsequently decays leptonically. The leptonic final states from the decay of the
₁₉₇₅ vector boson allow for leptonic triggering whilst at the same time significantly
₁₉₇₆ reducing the multijet background.

₁₉₇₇ A closely related analysis [123] has more recently measured the associated production
₁₉₇₈ of a Higgs boson decaying to b -quarks in events where the vector and Higgs bosons are
₁₉₇₉ highly boosted. The analysis is outlined in Section 7.1. Modelling studies performed
₁₉₈₀ by the author are detailed in Section 7.2, and the results of the analysis are presented

1981 in Section 7.4. The author contributed to various signal and background modelling
1982 studies, fit studies, and to the diboson unblinding effort. This analysis has been
1983 published in Ref. [123]. Figures and tables from Ref. [123] are reproduced here.

1984 **7.1 Analysis Overview**

1985 The boosted VH , $H \rightarrow b\bar{b}$ analysis is focused on the high transverse momentum
1986 regime, which has the benefit of being more sensitive to physics beyond the Standard
1987 Model [124], but the disadvantage of being more challenging due to the increased dif-
1988 ficulty in the accurate reconstruction of highly energy events (discussed in Chapter 4).
1989 In order to focus on the high- p_T regime, the reconstructed vector boson is required to
1990 have $p_T^V > 250$ GeV (see Section 7.1.2). Events are also split into two p_T^V bins with the
1991 first bin covering $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and the second covering $p_T^V > 400 \text{ GeV}$,
1992 which allows the analysis to account for the improved signal-to-background in the
1993 high- p_T regime.

1994 The previous ATLAS analysis in Ref. [105] was primarily sensitive to vector bosons
1995 with a more modest p_T^V boost in the region of 100–300 GeV. In this regime, the Higgs
1996 candidate was reconstructed using a pair of jets with radius parameter of $R = 0.4$,
1997 called small- R jets. However in the high- p_T regime, the decay products of the Higgs
1998 boson become increasingly collimated and the small- R jets may overlap. In order to
1999 avoid the associated problems and to aid in the reconstruction of the Higgs boson
2000 candidate, the present analysis uses instead a large- R jet with radius parameter
2001 $R = 1.0$ to reconstruct the Higgs boson candidate in all channels (see Section 3.4.3).
2002 The Higgs candidate is required to have exactly two ghost-associated and b -tagged
2003 variable-radius track-jets. The candidate large- R jet is reconstructed using jet
2004 substructure techniques, for example it is trimmed by removing soft and wide-angle
2005 components, which helps to remove particles from the underlying event and pileup
2006 collisions [125]. Refer to Section 3.4.3 for more details on jet reconstruction.

2007 On top of the binning in p_T^V , selected events are further categorised into 0-, 1- and
2008 2-lepton channels depending on the number of selected charged leptons (electrons
2009 and muons) are present in the reconstructed final state (also referred to as 0L, 1L,
2010 and 2L respectively). The 0-lepton channel targets the $ZH \rightarrow \nu\nu b\bar{b}$ process, the
2011 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 definition of the analysis regions. Signal enriched regions are marked with the label SR. There are regions with relatively large signal purity (HP SR) and with low purity (LP SR). Background enriched 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 [31] 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. [105]. 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. [74], 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 [60, 108, 160]. 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 [126] + GoSAM [128] + MnLO [129, 130]	NNPDF3.0NLO (*) [94]	Pythia 8.212 [96]	AZNLO	NNLO(QCD) + NLO(EW) [131–137]
$qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$	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 [138–142]
Top quark ($m_t = 172.5$ GeV)					
$t\bar{t}$	Powheg-Box v2 [126, 143]	NNPDF3.0NLO	Pythia 8.230	A14 [97]	NNLO+NNLL [144]
s-channel	Powheg-Box v2 [126, 145]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [146]
t-channel	Powheg-Box v2 [126, 145]	NNPDF3.0NLO	Pythia 8.230	A14	NLO [147]
Wt	Powheg-Box v2 [126, 148]	NNPDF3.0NLO	Pythia 8.230	A14	Approximate NNLO [149]
Vector boson + jets					
$W \rightarrow \ell\nu$	SHERPA 2.2.1 [150–153]	NNPDF3.0NNLO	SHERPA 2.2.1 [154, 155]	Default	NNLO [156]
$Z/\gamma^* \rightarrow \ell\ell$	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. 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 [157] using the internal algorithm in POWHEG-BOX v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for the $pp \rightarrow ZH$ process already includes the $gg \rightarrow ZH$ contribution. The $gg \rightarrow ZH$ process is normalised using the cross-section for the $pp \rightarrow ZH$ process, after subtracting the $gg \rightarrow ZH$ contribution. An additional scale factor is applied to the $gg \rightarrow VH$ processes as a function of the transverse momentum of the vector boson, to account for electroweak (EW) corrections at NLO. This makes use of the VH differential cross-section computed with HAWK [158, 159].

2061 in data [161–163]. The jet tagging strategy relies on extensive studies into track-jet
2062 b -tagging in boosted topologies [164, 165].

2063 The jet flavour labelling scheme is described in Section 3.4.3.

2064 7.1.3 Selection Criteria

2065 An extensive list of selection cuts are applied to each event in order to reject
2066 background events whilst retaining as many signal events as possible. A full list of
2067 selection cuts applied to the different analysis regions is given in Table 7.3, while
2068 some key selections are listed below.

2069 All channels are require events with at least one large- R jet with $p_T > 250 \text{ GeV}$
2070 and $|\eta| < 2.0$. The vector boson transverse momentum is also required to satisfy
2071 $p_T^V > 250 \text{ GeV}$. The Higgs candidate is chosen as the highest p_T large- R jet satisfying
2072 these requirements. As mentioned, the candidate large- R jet is required to have
2073 two ghost-assciated and b -tagged variable-radius track-jets. These track-jets are
2074 required to have at least two constituent tracks with $p_T > 500 \text{ MeV}$ and $|\eta| < 2.5$.
2075 The track-jets themselves must satisfy $p_T > 10 \text{ GeV}$ and $|\eta| < 2.5$.

2076 In the 0-lepton channel, trigger selections are applied using an E_T^{miss} trigger with a
2077 luminosity-dependent threshold between 70–110 GeV. In the 1-lepton electron sub-
2078 channel a combination of single electron triggers is used with minimum p_T thresholds
2079 between 24–26 GeV. In the muon sub-channel the same E_T^{miss} trigger as the 0-lepton
2080 channel is used. Since muons are not used for the E_T^{miss} trigger calculations, this
2081 is in effect a p_T requirement on the muon-neutrino system, which in the analysis
2082 phase space is more efficient than a single-muon trigger. The 2-lepton channel uses
2083 the same triggering strategy as the 1-lepton channel. In all channels, the trigger
2084 selections applied are fully efficient for events selected using the full requirements in
2085 Table 7.3.

2086 The combined selections in Table 7.3 result in a signal efficiency ranging from 6–16%
2087 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. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region.

2088 7.1.4 Control Regions

2089 The $t\bar{t}$ process presents a major background in the 0- and 1-lepton channels. In these
 2090 events, the Higgs candidate is often reconstructed from a correctly tagged b -jet from
 2091 the top decay $t \rightarrow Wb$, and an incorrectly tagged c - or light-jet from the subsequent
 2092 decay of the W , as shown in Fig. 7.1.

not sure
where the
0L ttbar
ETmiss
comes from

2093 The only known decay mode of the top quark is via the weak force to a W and
 2094 a down-type quark. (it is the only quark heavy enough to decay into an on-shell
 2095 W). Overwhelmingly (96% of the time) the down-type quark is a b -quark Hence, the
 2096 second top quark is also likely to result in a second tagged b -tagged track-jet outside
 2097 of the large- R Higgs candidate. To ensure sufficient $t\bar{t}$ rejections, 0- and 1-lepton
 2098 channel signal regions are defined using a veto on events with b -tagged track-jets
 2099 outside the Higgs-jet candidate. These events are used to construct a control region
 2100 (CR) which is enriched in $t\bar{t}$ events. The CR is used to constrain the normalisation
 2101 of the $t\bar{t}$ background in the fit.

2102 7.1.5 Background Composition

2103 After the selections described in Section 7.1.3 the number of background events
 2104 mimicking the VH , $H \rightarrow b\bar{b}$ signal is greatly reduced. However, the number of
 2105 background events still greatly outnumbers that of signal events. The background
 2106 processes are channel dependent. In the 0-lepton channel the dominant sources of
 2107 backgrounds are $Z+jets$ ($Z \rightarrow \nu\nu$) and $t\bar{t}$, with $W+jets$ and diboson events being
 2108 subdominant. In the event of $W \rightarrow \tau\nu$, and subsequent hadronic decay of the τ or
 2109 lack of successful reconstruction/selection of the leptonic decay products, $W+jets$
 2110 can also contribute to the 0-lepton channel. $t\bar{t}$ and $W+jets$ (with a leptonic decay
 2111 of the W as in $W \rightarrow \ell\nu$) are dominant in the 1-lepton channel, while single-top is
 2112 subdominant. In the 2-lepton channel, $Z+jets$ ($Z \rightarrow \ell\ell$) is again dominant followed
 2113 by $Z Z$ diboson events.

2114 The diboson background VV consists primarily of WZ and ZZ events in which the
 2115 Z decays to a pair of b -quarks. This process very closely matches the signal, with
 2116 a resonant peak occurring at $m_Z = 91$ GeV and so is considered as an irreducible
 2117 background ($V+b$ -jets is also irreducible).

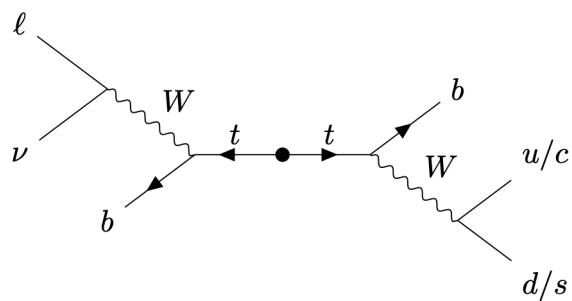
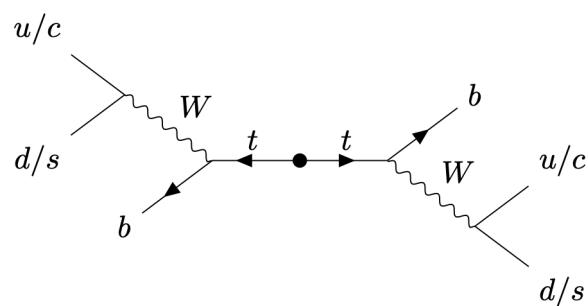
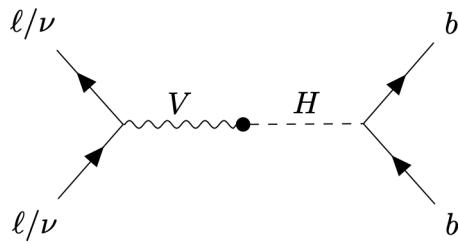


Figure 7.1: Diagrams of the signal process (top) and the 0-lepton and 1-lepton $t\bar{t}$ backgrounds (middle, bottom). Objects to the right of centre are reconstructed within the large- R jet. For the backgrounds, the large- R jet contains a mis-tagged c - or light-jet.

2118 The $t\bar{t}V$, $t\bar{t}H$ and multijet backgrounds are negligible in the analysis phase space
2119 after the selections have been applied, with the exception of the 1-lepton electron
2120 sub-channel, in which multijet background is not ignored. The multijet background
2121 is made up of jets with semileptonic heavy-flavour-hadron decays (e.g. $b \rightarrow c\ell\nu$) and
2122 jets which are mis-tagged by the flavour tagging algorithm MV2c10.

2123 The contributions from the different backgrounds are modelled using Monte Carlo
2124 event generators and the impacts on the analysis are studied in Section 7.2. The
2125 multijet background is not modelled but instead estimated using a data-driven
2126 technique.

2127 7.2 Systematic Uncertainties & Background 2128 Modelling

2129 Systemic uncertainties are extensively employed to give the fit model described in
2130 Section 7.3 enough flexibility to account for inaccuracies in the various inputs to the fit.
2131 Two main types of systematic uncertainty are considered: experimental and modelling.
2132 Experimental uncertainties arise due to the limited due to limited detector precision,
2133 imperfect reconstruction algorithms (in particular the b -tagging algorithms), and
2134 due to the imperfect measurement of pile-up and integrated luminosity. Modelling
2135 is the simulation processes relevant to the analysis using Monte-Carlo (MC) event
2136 generators, and is used to predict the outcome of the analysis. Modelling uncertainties
2137 arise due to the imperfections in the simulation of signal and background events, for
2138 example differences between event generators, or use of different model parameters
2139 when producing simulated events. In order to observe a certain process, for example
2140 VH , $H \rightarrow b\bar{b}$, an increase in the number of observed events with respect to the
2141 background-only hypothesis is looked for. The excess is often relatively small against
2142 the total number of background events, and hence accurate modelling of the expected
2143 number of background and signal events is crucial for successfully performing the
2144 analysis. Particular care is paid to the uncertainties on the modelling predictions as
2145 discussed in this section.

2146 Modelling uncertainties are described in detail in the following sections. Modelling
2147 uncertainties: *Nominal* samples are used as a reference to which different vari-

2148 ations can be compared. The nominal samples are chosen as the best possible
2149 representation of the underlying physical process. *Alternative* samples are used to
2150 understand inaccuracies that may be present in the nominal samples. Some aspect
2151 of the nominal model is varied, and the discrepancy with respect to the nominal
2152 model is quantified. The discrepancy is used to systematic uncertainty associated
2153 with the model parameter which was changed.

2154 Modelling studies involving c - and light-jets is hampered by the low available statistics
2155 of jets pass the analysis selections, due to the high rejection rates of the b -tagging
2156 algorithm MV2c10. For modelling studies therefore, truth tagging (TT) is employed
2157 to ensure sufficient numbers of jets are available to calculate uncertainties. TT works
2158 by computing a 2-dimensional efficiency map using the jet p_T and jet η . The two
2159 leading track-jets associated to the large- R jet automatically passes the b -tagging
2160 requirement, and are weighted based on their p_T and η using the pre-calculated
2161 efficiency map.

2162 7.2.1 Sources of Systematic Uncertainties

2163 This section briefly describes the different sources of uncertainty in the predictive
2164 model used in the analysis, and how each source of uncertainty is implemented within
2165 the analysis framework. Considered sources of systematic uncertainty are listed in
2166 Table 7.4. For each source of uncertainty, acceptance and shape uncertainties are
2167 derived.

2168 QCD Scales

2169 The $V+jets$ matrix element calculations contains infrared and ultraviolet divergences.
2170 These are handled by introducing spurious parameters corresponding to the renormalisation scale (μ_R) and factorisation scale (μ_F). Physical observables are not
2171 dependent on these parameters when using the infinite perturbation series expansion,
2172 however at some fixed order in QCD a limited dependence is present. To assess the
2173 impact of this, both μ_R and μ_F are independently varied from their nominal values
2174 by factors of 0.5 and 2 to account for higher order corrections to the calculation of
2175 the matrix element used to simulate the process.

2177 PDF Sets

2178 Parton distribution functions (PDFs) specify the probability of finding a parton
2179 with a given momentum inside a hadron (in this case, inside colliding protons).
2180 PDFs have to be derived from data and are a significant source of uncertainty in
2181 analyses of hadronic collision data. There are three sources of PDF uncertainties:
2182 the statistical and systematic errors on the underlying data used to derive the PDFs,
2183 the theory which is used to describe them (which is based on some fixed order
2184 perturbative QCD expansion), and finally the procedure which is used to extract the
2185 PDFs from the data. PDF-related uncertainties were derived following Ref. [157].
2186 This involves considering 100 PDF replicas which, when combined, form a central
2187 value and associated uncertainty, and also in parallel direct changes to the central
2188 values of PDFs using the MMHT2014 [166] and CT14NLO [167] PDF sets.

2189 Event Generator

2190 The choice of parton shower (PS) and underlying event (UE) generators can affect
2191 the analysis outcome. Changing these models modifies several aspects of the event
2192 generation at the same time, such as the accuracy of matrix element predictions and
2193 different approaches to parton showering. This change tends to lead to the largest
2194 discrepancy with respect to the nominal samples.

2195 Resummation and Merging Scales

2196 Resummation is a technique used in QCD to help cope with calculations involving
2197 disparate energy scales, and involves the introduction of an associated resummation
2198 scale, the choice of which introduces some systematic uncertainty into the model.
2199 Parton showering models are accurate when simulating low- p_T radiation, however
2200 inaccuracies start to arrive when simulating hard emissions. To combat this, par-
2201 ton showering models utilise more precise matrix element calculations above some
2202 momentum threshold. The choice of threshold, or *merging scale* introduces some
2203 uncertainty into the final result. Resummation (QSF) and merging (CKKW) scale
2204 variations are available for a subset of the SHERPA samples. The number of available
2205 events is significantly lower than the number of events in the nominal sample, and no

- 2206 statistically significant discrepancy with respect to the nominal samples is observed.
2207 The corresponding uncertainties and therefore neglected.

2208 7.2.2 Implementation of Variations

- 2209 Modelling variations are implemented in different ways, depending on the associated
2210 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.

- 2211
- 2212 is computationally expensive, a technique in state of the art simulation packages
2213 is to store some sources of variation as internal weights, which can be generated
2214 alongside the nominal samples, saving computation time. The nominal sample then
2215 effectively contains information about an ensemble of different samples, corresponding
2216 to different model parameters, which are accessible via reweightings. When filling
2217 histograms for the variations, bins are incremented by the internal weight of the
2218 event associated with the variation in question.

- 2219 While the inclusion of internal weight variation in MC event generators has decreased
2220 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently
2221 some sources of systematic uncertainty that are unable to be stored as internal weight
2222 variations due to technical limitations. Two examples are the choice of resummation
2223 and merging scales. A method to parameterise the systematic variation using
2224 one sample, and to then apply this parameterisation to another sample, has been

2225 developed by ATLAS [168]. This method was used to derive resummation and
2226 merging uncertainties for the nominal SHERPA 2.2.1 sample, using a previous (lower
2227 statistic) SHERPA 2.1 alternative sample. The resulting uncertainties were studied
2228 and found to be negligible in comparison with systematics from other sources.

2229 **7.2.3 Vector Boson + Jets Modelling**

2230 After event selection, the $V+jets$ background is a dominant background in all three
2231 analysis channels as described in Section 7.1.5. The $V+jets$ samples are split into
2232 categories depending on the truth flavour of the track-jets which are ghost-associated
2233 to the large- R jet Higgs candidate. The categories are $V+bb$, $V+bc$, $V+bl$, $V+cc$,
2234 $V+cl$, $V+ll$, and $V+hf$ refers collectively to the categories containing at least one
2235 b - or c -jet. $V+bb$ is dominant generally accounting for 80% of the jets, while $V+hf$
2236 accounts for around 90% of jets. The full flavour composition breakdown for each
2237 channel and analysis region are given in Tables 7.5, 7.7 and 7.8.

2238 In order to access uncertainties associated with the use of MC generators, variations
2239 of the data are produced using alternative generators or variation of nominal
2240 generator parameters as described in Section 7.2.2. As described in Section 7.1.1,
2241 the nominal MC event generator used for $V+jets$ events is SHERPA 2.2.1, while
2242 MADGRAPH5_AMC@NLO+PYTHIA8 (which uses a different parton showering
2243 model) is used as an alternative generator.

2244 Modelling systematics can have several impacts, including affecting the overall
2245 normalisation for different processes, and the relative acceptances between different
2246 analysis regions (i.e. migrations between HP and LP SRs, between the SR and CR,
2247 and between p_T^V bins), and the shapes of the m_J distributions. Since the fit model
2248 fits only the large- R jet mass m_J to data, all shape uncertainties are estimated with
2249 respect to this observable. Several sources of uncertainty, summarised in Section 7.2.1,
2250 have been assessed.

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

Table 7.5: 0-lepton $W+jets$ nominal sample flavour composition and total event yield.

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

Table 7.6: 1-lepton $W+jets$ nominal sample flavour composition and total event yield.

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

Table 7.7: 0-lepton $Z+jets$ nominal sample flavour composition and total event yield.

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

Table 7.8: 2-lepton $Z+jets$ nominal sample flavour composition and total event yield.

2251 Acceptance Uncertainties

2252 Several different types of acceptance uncertainties have been calculated and imple-
2253 mented as nuisance parameters in the fit. These account for uncertainty in the overall
2254 number of events in each channel, and for the migration of events between different
2255 analysis regions. The acceptance uncertainties relevant to the $V+jets$ processes are
2256 summarised below.

- 2257 • Overall normalisation:** only relevant where normalisation cannot be left
2258 floating (determined as part of the fit). The $V+hf$ component is left floating in
2259 the fit. For other components, independent normalisations used for $W+hf$ and
2260 $Z+hf$. The contributions are mainly determined by the 1-lepton (for $W+hf$)
2261 and 2-lepton (for $Z+hf$) SRs respectively and then extrapolated to 0-lepton
2262 channel.

- 2263 • SR-to-CR relative acceptance:** the uncertainty on the normalisation of the
2264 signal region due to events migrating between the signal and control regions.

- 2265 • HP-to-LP relative acceptance:** the uncertainty on the normalisation of the
2266 high-purity (HP) signal region due to events migrating between the high- and
2267 low-purity signal regions.

- 2268 • Medium-to-high p_T^V relative acceptance:** describes any shape effect in p_T^V
2269 distribution, given that the analysis only uses two p_T^V bins (medium and high).

- 2270 • Flavour relative acceptance:** for each flavour $V+xx$, where $xx \in \{bc, bl, cc\}$
2271 the ratio of $V+xx/V+bb$ events is calculated. This corresponds to the uncer-
2272 tainty of Vbb events due to the miss-tagging of other flavours Vxx .

- 2273 • Channel relative acceptance:** corresponding to the uncertainty in the nor-
2274 malisation of $V+jets$ events events due to the migration of events between
2275 channels.

2276 The uncertainties arising from the different sources described in Section 7.2.1 are
2277 summed in quadrature to give a total uncertainty on each region. A summary of the
2278 different acceptance uncertainties that were derived in this way and subsequently
2279 applied in the fit are given in Table 7.9. An effort has been made, wherever possible,
2280 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. 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.

2281 Shape Uncertainties

2282 In order to derive shape uncertainties (which as the name suggests affect shapes but
 2283 not overall normalisations of distributions), the following procedure is carried out.
 2284 Normalised distributions of the reconstructed large- R Higgs candidate jet mass m_J
 2285 are compared for the nominal sample and variations. For each variation, the ratio of
 2286 the variation to nominal is calculated, the up and down variations are symmetrised,
 2287 and an analytic function is fit to the symmetrised ratio. If different analysis regions
 2288 or channels show the same pattern of variation, a common uncertainty is assigned.
 2289 An example of a significant source of uncertainty, arising from choice of factorisation
 2290 scale μ_R is shown in Fig. 7.2. HP SRs split into medium and high p_T^V bins are shown
 2291 for the 0-lepton channel for $W+\text{hf}$ and $Z+\text{hf}$ jets. The 0- and 1-lepton channels for
 2292 the $W+\text{hf}$ contribution and the 0- and 2-lepton channels for the $Z+\text{jets}$ contribution
 2293 are merged, since the shapes in m_J are consistent across channels. An exponential
 2294 function $e^{p_0 + p_1 x} + p_2$ has been fitted to the ratio of the normalised distributions.
 2295 The magnitude of the variation does show p_T^V dependence, and so two separate
 2296 uncertainties are added in the fit, and applied individually in each p_T^V region.
 2297 The shape uncertainties for μ_R were derived on the SRs but are also applied to the
 2298 CRs, as the low statistics in the CRs make it difficult to derive dedicated shape
 2299 uncertainties. All the shape uncertainties are fully correlated accross regions.
 2300 A comparison of the m_J shapes between SHERPA and MADGRAPH is shown in
 2301 Fig. 7.3. The plots are split by process and channel, but merged in SR purity and p_T^V
 2302 bins reflecting similarities between the m_J shapes across these regions. Due to the low
 2303 statistics available for the alternate MADGRAPH sample, and the lack of statistically
 2304 significant variation between the samples, no associated shape uncertainty is added
 2305 to the fit in this case.
 2306 The impacts of variations in the factorisation scale μ_F and the choice of PDF set on
 2307 m_J shape were also found to be negligible in comparison with μ_R and are hence not
 2308 associated uncertainty was added to the fit.

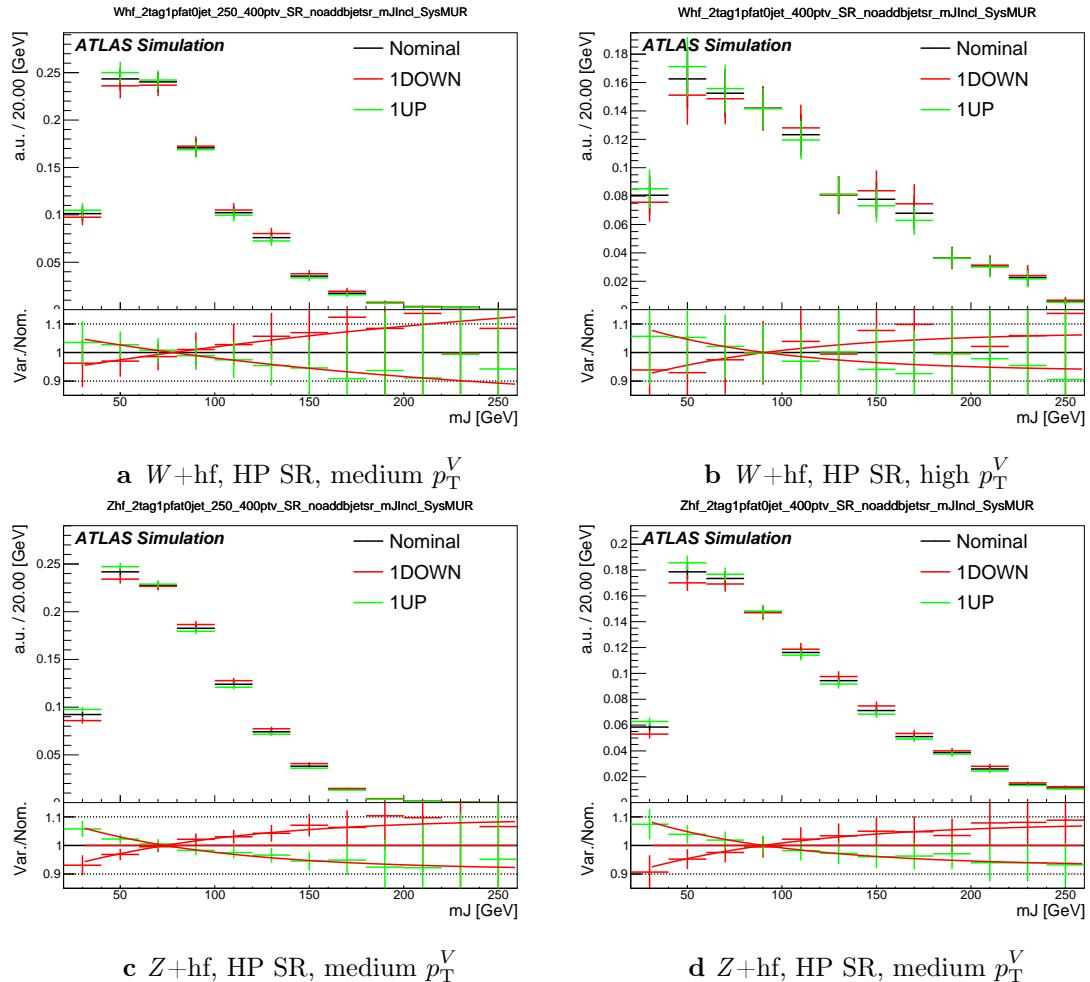


Figure 7.2: Normalised leading large- R jet mass distribution from Z and $W + hf$ processes in the HP SR of the 0-lepton channel. The renormalisation scale μ_r has been varied by a factor of 2 (1up) and 0.5 (1down). An exponential function is fitted to the ratio between the nominal and variation samples.

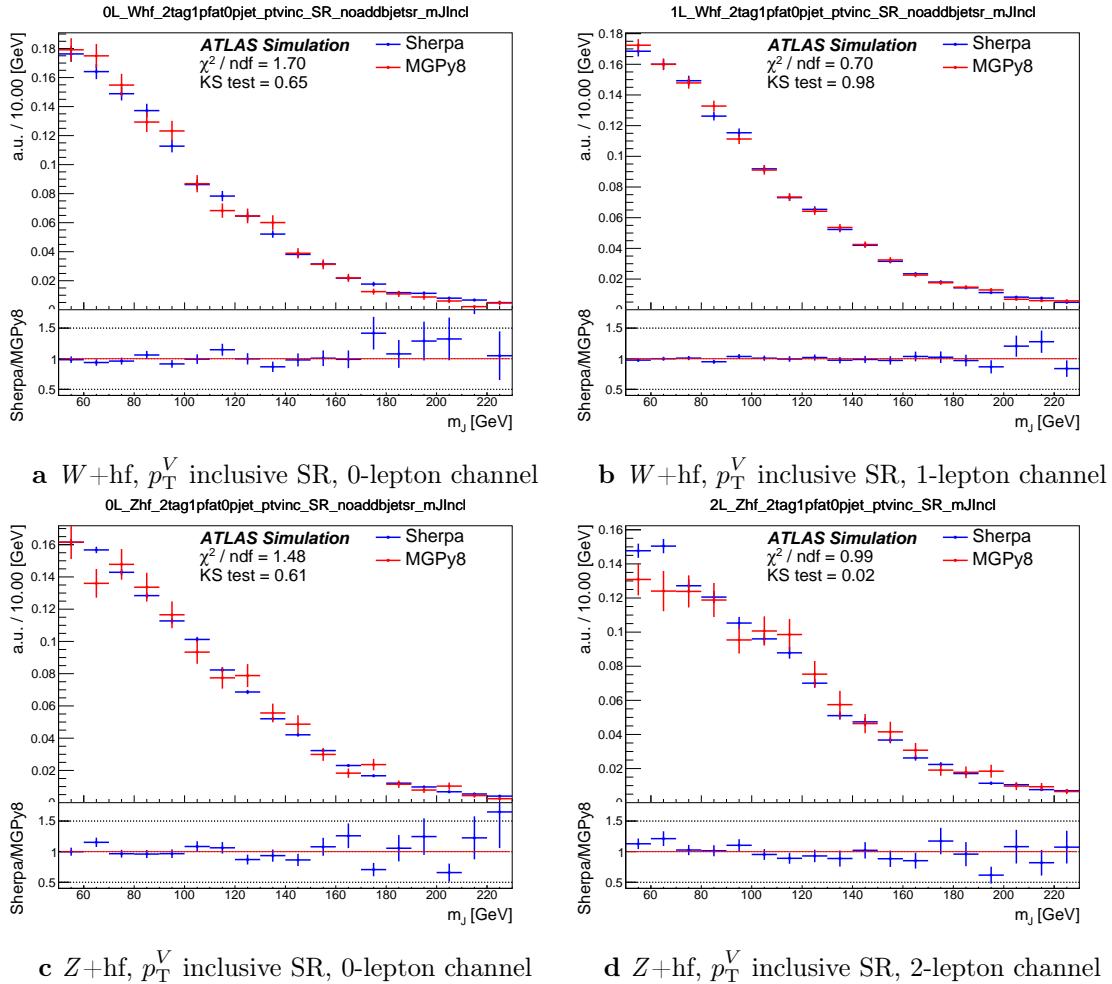


Figure 7.3: The comparison on m_J shapes between SHERPA and MADGRAPH samples from $V + \text{hf}$ process in p_T^V inclusive signal regions. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.

2309 7.2.4 Diboson Modelling

2310 The uncertainties for the diboson background generally follows that of $V+jets$. How-
2311 ever an alternative sample was generated using POWHEG interfaced with PYTHIA8,
2312 using the AZNLO shower tune with the CTEQ6L1 PDFs [169]. Unlike SHERPA,
2313 POWHEG models the off-shell Z contribution at NLO.

2314 Acceptance and shape uncertainties are derived in an analogous fashion to $V+jets$.

2315 Acceptance Uncertainties

2316 Diboson acceptance uncertainties are summarised in Table 7.10. Variations from μ_R ,
2317 μ_F , PDF choice and alternative generator are considered and are combined combined
2318 through a sum in quadrature as described in Section 7.2.3. The largest modification
2319 to the nominal acceptance results from the POWHEG+PYTHIA8 alternate sample,
2320 which modifies several parts of the generative model at the same time. Since the
2321 diboson contribution to the $t\bar{t}$ control region is small, no SR-to-CR relative acceptance
2322 uncertainty is required.

2323 For the WZ contribution, uncertainties are derived using the 1-lepton channel and
2324 applied to all three channels. An additional 8% channel migration uncertainty
2325 is applied on the 0-lepton channel. For the ZZ contribution, the normalisation
2326 uncertainty is calculated using the 2-lepton channel and applied to all three channels.
2327 The 0- and 1-lepton channels have a similar HP-to-LP relative acceptance uncertainty
2328 of 18%. The 1-lepton medium-to-high p_T^V relative acceptance is based off the value
2329 obtained from the 2-lepton channel. 30% and 18% channel migration uncertainties
2330 are applied in the 0- and 1-lepton channels respectively.

2331 Since the contribution from WW is small, dedicated studies are not performed, but
2332 a 25% normalisation uncertainty is applied in all the three channels which is based
2333 on the modelling studies performed for the previous analysis [105].

2334 Shape Uncertainties

2335 Diboson shape uncertainties are derived in a similar fashion to $V+jets$. Only the
2336 uncertainties associated with systematic variation of μ_R and the event generator

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. All uncertainties except channel extrapolation uncertainties are fully correlated between ZZ and WZ processes and channels.

have a non-negligible impact on the m_J shape. Variation of μ_R produces consistent m_J shape impacts across all regions and channels, and hence only a single associated uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted to the symmetrised ratio.

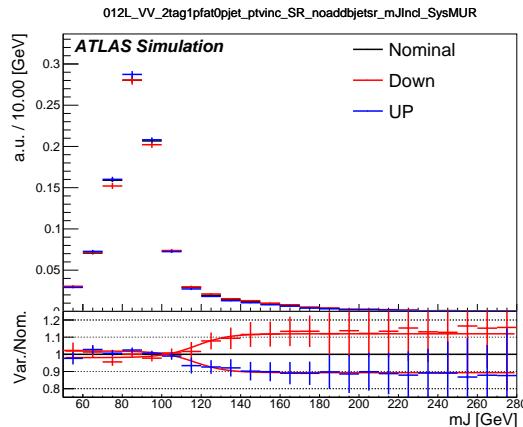


Figure 7.4: Normalised leading large- R jet mass distribution from WZ and ZZ process, merged among all the signal regions and lepton channels. The renormalisation scale μ_R has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function.

The comparison between the nominal SHERPA and alternative POWHEG+PYTHIA samples is shown in Fig. 7.5 for the 0- and 1-lepton channels for both WZ and ZZ processes. For these channels, the shape of m_J varies in opposite directions in the LP and HP signal regions. Shapes are similar between p_T^V bins, the 0- and 1-lepton channels and for WZ and ZZ . A third order polynomial is fitted to the ratio, and this function transitions to a constant piecewise function in the high

mass region to accurately represent the shape taking into account large statistical uncertainties. Dependence on event generator was found to be negligible within statistical uncertainty in the 2-lepton channel. All diboson shape uncertainties are fully correlated in the fit.

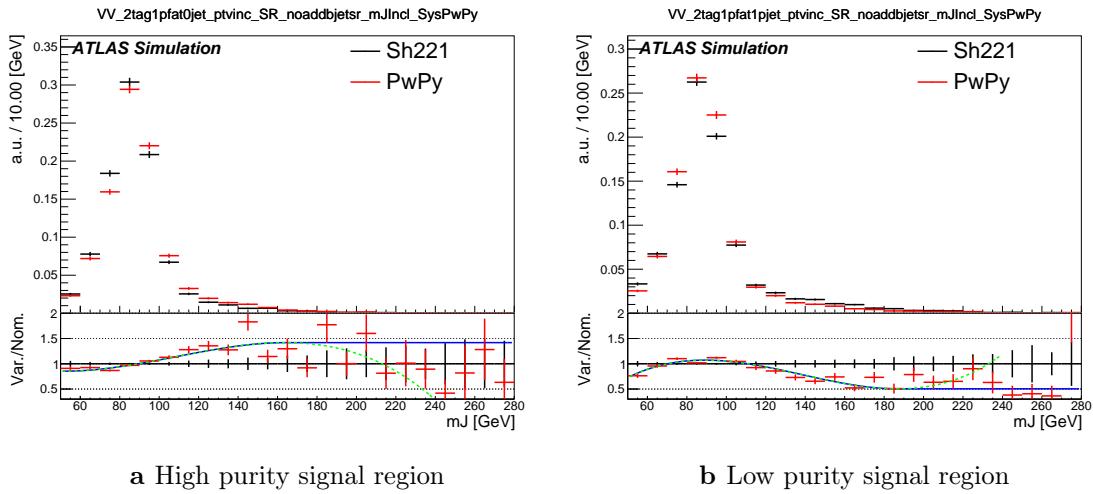


Figure 7.5: The comparison on m_J shapes between SHERPA and POWHEG+PYTHIA 8 samples from WZ and ZZ process in high and low purity signal regions. p_T^V regions and 0- and 1-lepton channels are merged. The dashed green line shows the fitted third order polynomial function and the blue lines show the function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

7.3 Statistical Treatment

Selected events are used to perform a statistical test of the background-only hypothesis, i.e. a model which does not include the VH , $H \rightarrow b\bar{b}$ process. The test involves a binned global maximum-profile-likelihood fit from the model to the data using the m_J distribution, and combines all the analysis regions defined in Table 7.1. The test is based on the profile likelihood ratio test statistic. The signal strength $\mu = \sigma/\sigma_{SM}$ is defined as the ratio between the observed and predicted cross-sections, where $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ corresponds to the SM prediction. It is a parameter of interest (POI) which acts to scale the total number of signal events.

The present analysis makes use of two POIs. The first, μ_{VH}^{bb} , is the signal strength for the VH , $H \rightarrow b\bar{b}$ process, the primary process under investigation. The diboson

2363 production strength μ_{VZ}^{bb} for the VZ , $Z \rightarrow b\bar{b}$ process is measured simultaneously
 2364 and provides a validation of the analysis apparatus used for the primary $H \rightarrow b\bar{b}$
 2365 measurement. Alongside the two POIs, the predictive model depends on several
 2366 uninteresting parameters which are not the primary target of measurement. These
 2367 parameters are called nuisance parameters (NPs), collectively referred to as θ . Freely
 2368 floating background normalisations are implemented as NPs and are also extracted
 2369 during the fitting processes.

2370 7.3.1 Likelihood Function

2371 The statistical setup treats each bin as a Poisson counting experiment and is based on
 2372 the ROOSTATS framework [170]. The combined likelihood over N bins is constructed
 2373 as the product of Poisson probabilities in each bin. Considering the simplified case
 2374 of a single signal strength parameter μ , and neglecting sources of systematic or
 2375 statistical uncertainty, this is given by

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

2376 where s_i (b_i) is the expected number of signal (background) events in bin i , and n_i is
 2377 the number of observed data events in bin i .

2378 Treatment of Uncertainties

2379 Systematic uncertainties can modify the predicted signal and background yields s_i
 2380 and b_i . Each source of systematic uncertainty is taken into account by adding an
 2381 additional NP θ_j to the likelihood in the form of a Gaussian cost function. The
 2382 combined effect of the NPs is then

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

2383 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
 2384 value, and σ_j is the corresponding associated prior uncertainty on θ_j . As the fitted
 2385 value of the θ_j deviates from its nominal value, a cost is introduced. The presence

2386 of NPs modifies the likelihood as

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

2387 The predicted signal and background yields are also modified by the presence of the
2388 NPs with

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

2389 For NPs which are left freely floating in the fit, no corresponding Gaussian constraint
2390 is added to the likelihood.

2391 Statistical uncertainty is also present, and implemented using a dedicated NP for
2392 each bin which can scale the background yield in that bin. Statistical NPs are also
2393 implemented using a Gaussian constraint.

2394 Smoothing and Pruning

2395 Systematic uncertainties are smoothed and pruned in the fit. Smoothing accounts
2396 for the large statistical uncertainty present in some bins that are subject to large
2397 fluctuations. The smoothing procedure relies on the assumption that the impact of
2398 systematics should be approximately monotonic and correlated between neighbouring
2399 bins.

2400 In addition to smoothing, pruning is the process of removing from the fit those
2401 systematics which only have a very small effect. This improves the stability of the
2402 fit by reducing the number of degrees of freedom. Acceptance uncertainties are
2403 pruned in a given region if they have a variation of less than 0.5%, or if the up and
2404 down variations have the same sign in that region. Shape uncertainties are pruned
2405 in a given region if the deviation in each bin is less than 0.5% in that region. In
2406 addition, acceptance and shape uncertainties are neglected in a given region for any
2407 background which makes up less than 2% of the total background in a given region.

2408 Fit Procedure and Statistical Tests

2409 The best-fit value of μ , denoted $\hat{\mu}$, is obtained via an unconditional maximisation
 2410 of the likelihood. The likelihood is also used to construct a statistical test which can
 2411 confirm or reject the background-only hypothesis. The test statistic q_μ is constructed
 2412 from the profile likelihood ratio, as in

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

2413 where $\hat{\mu}$ and $\hat{\theta}$ are chosen to maximise the likelihood \mathcal{L} , and the profile value $\hat{\theta}_\mu$ is
 2414 obtained from a conditional maximisation fo the likelihood for a specific choice of
 2415 $\mu = 0$ corresponding to the background-only hypothesis.

2416 The test statistic is used to construct a p -value which is used to confirm or accept
 2417 the background-only hypothesis. The p -value is typically reported in terms of the
 2418 significance Z , defined as the number of standard deviations for a Gaussian Normal
 2419 distribution which will produce a one-sided tail integral equal to the p -value, as in

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

2420 Typically a value of $Z = 3$ constitutes *evidence* of a processes, while $Z = 5$ is required
 2421 for a *discovery*. Alongside the p -value, the best-fit value of the signal strength $\hat{\mu}$ and
 2422 its corresponding uncertainty are typically quoted, and compared to their expected
 2423 values (see Section 7.3.3).

2424 7.3.2 Background Normalisations

2425 The normalisation of the largest backgrounds are left floating and are determined in
 2426 the fit. The corresponding postfit background normalisations are listed in Table 7.11.
 2427 A single normalisation factor is used for $W+\text{hf}$ and $Z+\text{hf}$, which constitue more
 2428 than 90% of the total $V+\text{jets}$ background, since the use of independent factors in
 2429 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. The errors represent the combined statistical and systematic uncertainties.

2430 The normalisations and shapes of all other backgrounds, with the exception of the
2431 multijet background which is estimated using a data driven technique, are initialised
2432 using the simulated samples.

2433 7.3.3 Asimov Dataset & Expected Results

2434 The Asimov dataset is constructed by replacing the data with the sum of the signal
2435 and background predictions $n_i = s_i + b_i$. A fit to this dataset using the nominal
2436 values of the NPs from the simulation will recover the input values and is useful for
2437 studying constraints on and correlations between the NPs.

2438 Alternatively, a conditional fit to the Asimov dataset can be performed using values
2439 of the background NPs which are determined from an unconditional fit to data. The
2440 signal NPs and POIs are fixed at their nominal values from the SM simulation. The
2441 result of this fit can be used to calculate expected (median) significances, which can
2442 be compared to their observed values.

2443 7.4 Results

2444 In the present analysis, the two signal strength parameters μ_{VH}^{bb} and μ_{VZ}^{bb} are extracted
2445 from a simultaneous maximisation of the likelihood described in Section 7.3. The
2446 results of the analysis are summarised in this section. Post-fit m_J distributions are
2447 shown in Section 7.4.1. The observed signal strengths are given in Section 7.4.2,
2448 along with observed and expected significances. Finally in Section 7.4.3 the impact
2449 of systematic uncertainties on the results is examined.

2450 7.4.1 Post-fit Distributions

2451 In addition to the observed significance and signal strength, it is also useful to study
 2452 the post-fit m_J distributions to compare the simulation and data using the best-fit
 2453 values $\hat{\mu}$ and $\hat{\theta}$. Post-fit m_J distributions are given for the signal regions in the 0-,
 2454 1- and 2-lepton channels in Fig. 7.6. The LP and HP regions are merged for the
 2455 0- and 1-lepton channels. The plots show large falling backgrounds, predominantly
 2456 made up of W -jets and Z -jets events, and a signal distribution corresponding to
 2457 the Standard Model Higgs boson peaking around $m_H = 125$ GeV. In general there is
 2458 a good level of agreement between the simulation and data, indicating the fit model
 2459 is performing as expected. Fig. 7.7 shows the post-fit plots for the $t\bar{t}$ control regions.
 2460 Again, a good level of agreement is observed given the statistical uncertainties on
 2461 the distributions.

2462 7.4.2 Observed Signal Strength & Significance

2463 The measured signal strength is computed as the ratio between the measured signal
 2464 yield to the prediction from the SM. The combined result for all three lepton channels
 2465 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).
 2466 Both results include a full breakdown of the systematic and statistical uncertainties.

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

2467

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

2468 The results for μ_{VH}^{bb} and μ_{VZ}^{bb} agree with the expectation from the SM within their
 2469 combined uncertainty. The μ_{VH}^{bb} measurement is dominated by statistical uncertainty,
 2470 while the μ_{VZ}^{bb} measurement is dominated by systematic sources of uncertainty. These
 2471 measured signal strength for μ_{VZ}^{bb} corresponds to an observed significance of 2.1
 2472 standard deviations, with an expected (median) significance given the SM prediction
 2473 of 2.7 standard deviations obtained using the method described in Section 7.3.3. The
 2474 diboson observed (expected) signal strength significance is 5.4 (5.7). These results
 2475 are summarised in Fig. 7.8, which shows the background-subtracted m_J distribution.
 2476 A clear signal excess is visible around the Higgs mass of $m_H = 125$ GeV.

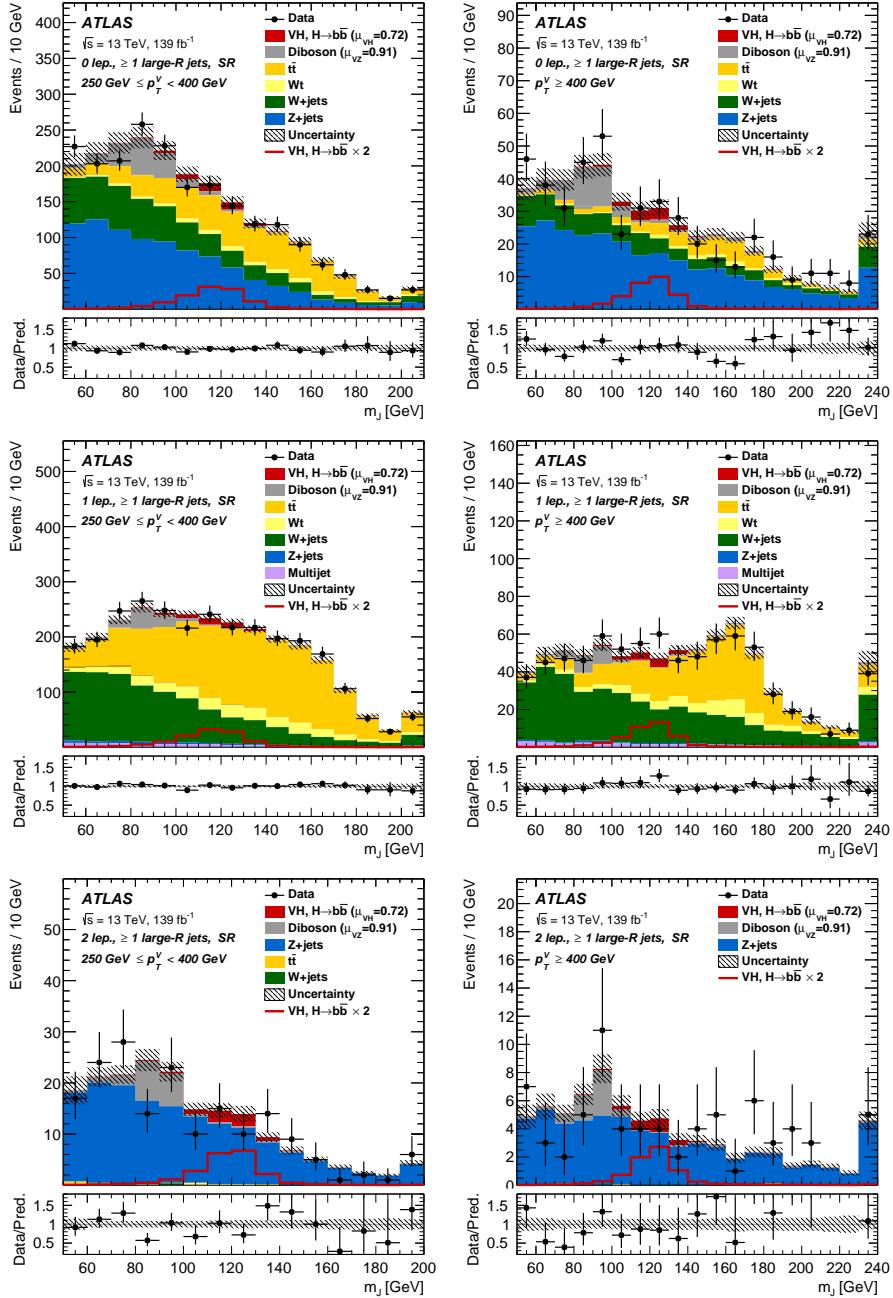


Figure 7.6: The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$. The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions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

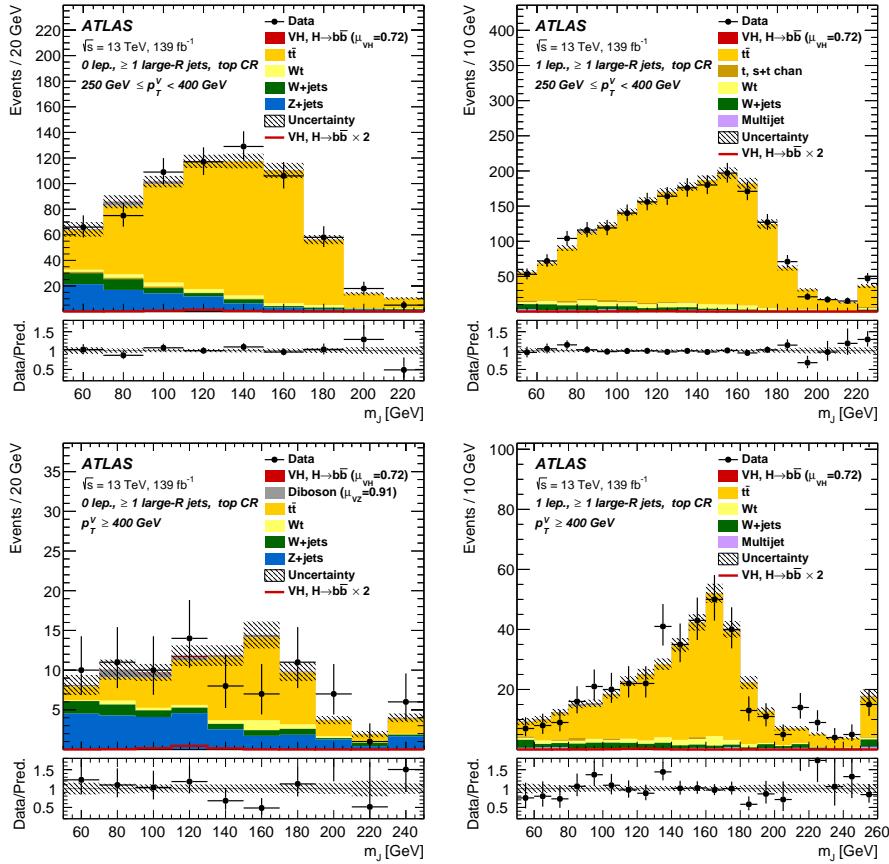


Figure 7.7: The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. The background contributions after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

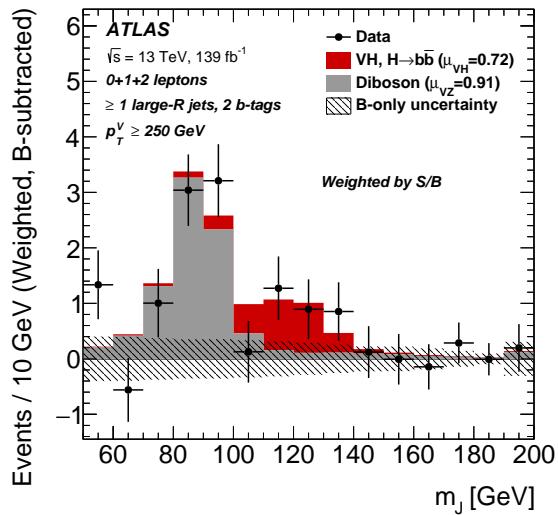


Figure 7.8: m_J distribution in data after subtraction of all backgrounds except for the WZ and ZZ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected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

²⁴⁷⁷ **Compatability Studies**

²⁴⁷⁸ Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting μ_{VH}^{bb}
²⁴⁷⁹ into three separate POIs, one for each channel. A simultaneous fit to the channel
²⁴⁸⁰ specific signal strengths can then be performned, which allows a comparison of the
²⁴⁸¹ contributions from each channel. Fig. 7.9 compares the best-fit signal strengths.
²⁴⁸² The 0- and 1-lepton channels show a signal strength which is consistent with the
²⁴⁸³ SM prediction, while the 2-lepton channel shows a small deviation within the 1σ
²⁴⁸⁴ uncertainty. Overall, good compatibility is observed via a χ^2 test with a corresponding
²⁴⁸⁵ p -value of 49%.

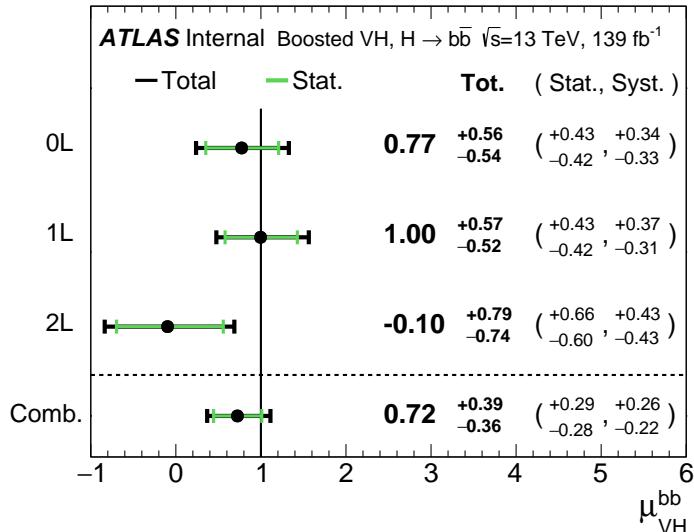


Figure 7.9: Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels splitted) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a χ^2 difference test and results in a p -value of 49%.

²⁴⁸⁶ **7.4.3 Impact of Systematics**

²⁴⁸⁷ The impact of systematic uncertainties on the final fitted value $\hat{\mu}_{VH}^{bb}$ is studied looking
²⁴⁸⁸ at the NP rankings, and the uncertainty breakdown.

²⁴⁸⁹ Fig. 7.10 shows the NP ranking, which is used to visualise which out of the many
²⁴⁹⁰ NPs involved in the fit have the largest impact on the sensitivity to the fitted POI.
²⁴⁹¹ To obtain the ranking, a likelihood scan is performed for each NP θ_j . First, an

2492 unconditional fit is used to determine $\hat{\theta}_j$. From this best-fit point, the NP is varied
2493 in steps and the likelihood is recomputed until the $\pm 1\sigma_{\hat{\theta}_j}$ values are reached. For
2494 each corresponding value of θ_j , the change in the best-fit value of the POI, $\Delta\hat{\mu}_{VH}^{bb}$
2495 is calculated and used to rank the NPs. Also shown in Fig. 7.10 are the pulls and
2496 constraints for the highest ranked NPs.

2497 The experimental uncertainty on the signal large- R jet mass resolution (JMR) has
2498 the largest impact of any NP. It is a significant contributor to the overall uncertainty
2499 on μ_{VH}^{bb} in Eq. (7.7). JMR and jet energy scale (JES) uncertainties also have impacts
2500 for the $V+jets$ background and for the diboson background. The freely-floating
2501 $Z+hf$ normalisation is the second highest ranked NP, and is heavily constrained by
2502 the fit. The VZ POI μ_{VZ}^{bb} is also a significant NP when considering the primary μ_{VH}^{bb}
2503 measurement.

2504 The NP ranking highlights individual NPs which have a large impact on the POI
2505 measurement sensitivity. Complementary information is provided at a higher level
2506 by considering the overall impact of different groups of systematics. The groups
2507 are constructed from NPs which have similar physical origin. The impact on each
2508 group is calculated by running a fit with all the NPs in the given group fixed to their
2509 nominal values. The uncertainty on the POI extracted from this fit is subtracted
2510 in quadrature from the the uncertainty on the POI from the nominal fit, and the
2511 resulting values are provided as the impact for each group. The full breakdown
2512 for the observed impact of uncertainties on the μ_{VH}^{bb} signal strength is provided in
2513 Table 7.12. The total systematic impact is the difference in quadrature between the
2514 nominal uncertainty on μ_{VH}^{bb} and the combined statistical impact. The “data stat
2515 only” group fixes all NPs at their nominal value, while the total statistical impact
2516 fixes all NPs except floating normalisations. The floating normalisations group fixes
2517 only the NPs associated with normalisation which are left floating in the fit. The
2518 uncertainty on μ_{VH}^{bb} is dominated by combined statistcal effects (0.28), although the
2519 combined impact of systematics (0.24) is of a comparable size. The signal largest
2520 group is the data stat uncertainty (0.25), demonstrating that the analysis would
2521 benefit from an increased integrated luminosity or improved efficiency to select signal
2522 relevant events (recall from Section 7.1.3 the signal efficiency is in the range of
2523 10%). Of the experimental systematic sources of uncertainty, the dominant impact
2524 is the experimental uncertainties associated with the reconstruction of large- R jets
2525 (0.13). Other experimental sources of uncertainty are small in comparison. Modelling

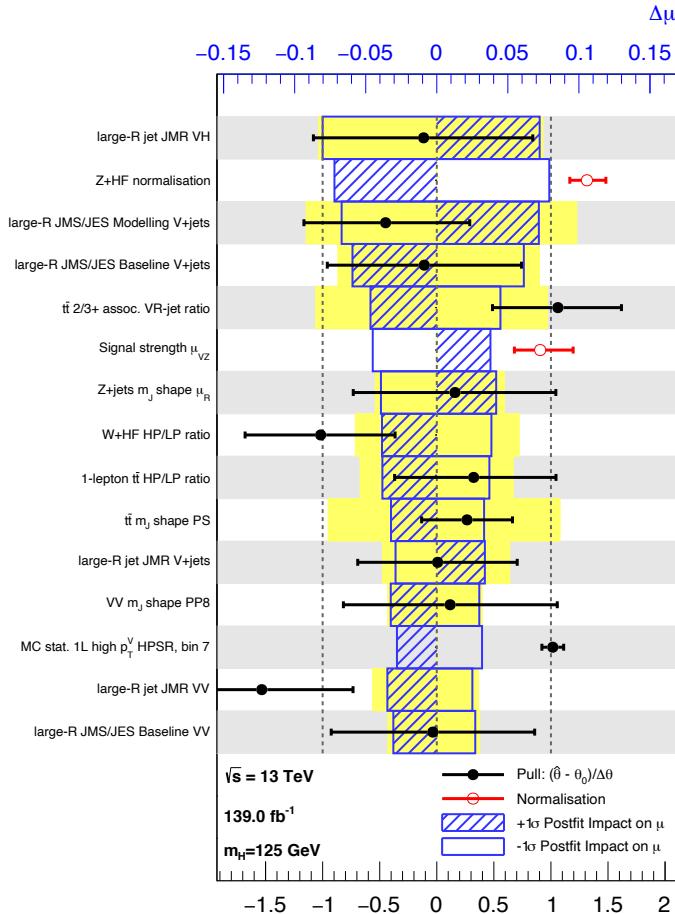


Figure 7.10: Impact of systematic uncertainties on the fitted VH signal-strength parameter $\hat{\mu}_{VH}^{bb}$ sorted in decreasing order. The boxes show the variations of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual nuisance parameter to its post-fit value modified upwards or downwards by its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of up- and down-variations can be distinguished via the dashed and plane box fillings. The yellow boxes show the pre-fit impact (top x -axis) by varying each nuisance parameter by ± 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.

2526 uncertainties also have a large contribution to the overall systematic uncertainty. The
2527 biggest contribution to the overall uncertainty is the combined statistical uncertainty
2528 on the simulated samples (0.09), which contain only a finite number of events. Out of
2529 the backgrounds, the $W+jets$ and $Z+jets$ have the highest (0.06) and second-highest
2530 (0.05) impact respectively.

2531 7.5 Improved b -tagging using GNNs

2532 7.6 Conclusion

2533 Work has been carried out as part of the boosted VHbb analysis group to understand,
2534 and implement in the global profile likelihood fit, systematic uncertainties on $V+jets$
2535 samples. This background modelling work is an essential part of the success of
2536 the analysis. So far the fit has proved stable with the inclusion of the $V+jets$
2537 uncertainties, and detailed studies are now underway to determine the causes behind
2538 any observed pulls of the added NPs. Additional work is ongoing to help with the
2539 derivation of uncertainties on diboson samples, another important background. The
2540 analysis is already advanced, and is now progressing into its final stages. Publication
2541 is expected in the new year.

2542 This analysis would benefit greatly from the improved high p_T b -tagging enabled by
2543 GN1.

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. 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.

2544 Chapter 8

2545 Conclusion

2546 flavour tagging at high pT is important because

2547 tracking suffers from several problems which might be improved by

2548 algorithmic gains in flavour tagging can improve things

2549 ultimately analyses will benefit greatly

2550 future work:

2551 • improve tracking

2552 • use more info for flavour tagging

2553 • improve jet labelling

2554 • boosted xbb tagging

2555 • lead to dihiggs

2556 **Bibliography**

- 2557 [1] A. Buckley, *A class for typesetting academic theses*, (2010). [http://ctan.tug.org/
2558 tex-archive/macros/latex/contrib/heptesis/heptesis.pdf](http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf).
- 2559 [2] L. Morel, Z. Yao, P. Cladé and S. Guellati-Khélifa, *Determination of the
2560 fine-structure constant with an accuracy of 81 parts per trillion*, *Nature* **588** (2020)
2561 pp. 61–65.
- 2562 [3] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al.,
2563 *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* **606**
2564 (2022) pp. 479–483.
- 2565 [4] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev.
2566 Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- 2567 [5] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995)
2568 pp. 2632–2637 [[hep-ex/9503003](#)].
- 2569 [6] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV
2570 Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) pp. 252–255.
- 2571 [7] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy
2572 Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* **122**
2573 (1983) pp. 103–116.
- 2574 [8] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504**
2575 (2001) pp. 218–224 [[hep-ex/0012035](#)].
- 2576 [9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
2577 *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- 2578 [10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.*
2579 **13** (1964) pp. 508–509.
- 2580 [11] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*

- 2581 *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 2582 [12] ATLAS Collaboration, *Observation of a new particle in the search for the Standard
2583 Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)
2584 p. 1 [[1207.7214](#)].
- 2585 [13] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS
2586 experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 2587 [14] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)
2588 p. 083C01.
- 2589 [15] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge
2590 invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 2591 [16] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)
2592 pp. 579–588.
- 2593 [17] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 2594 [18] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel
2595 symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**
2596 (1968) pp. 367–377.
- 2597 [19] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,
2598 *Phys. Rev.* **104** (1956) pp. 254–258.
- 2599 [20] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental
2600 test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 2601 [21] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of
2602 conservation of parity and charge conjugation in meson decays: the magnetic moment
2603 of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 2604 [22] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,
2605 *Phys. Rev.* **117** (1960) pp. 648–663.
- 2606 [23] J. Goldstone, *Field theories with «superconductor» solutions*, *Il Nuovo Cimento
2607 (1955-1965)* **19** (1961) pp. 154–164.
- 2608 [24] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs
2609 Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](#).
- 2610 [25] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*

- 2611 *ATLAS detector*, ATLAS-CONF-2018-036 (2018).
- 2612 [26] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys. Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](https://arxiv.org/abs/1808.08242)].
- 2614 [27] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 2615 [28] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) pp. S08002–S08002.
- 2617 [29] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) p. S08004.
- 2619 [30] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of Instrumentation* **3** (2008) pp. S08005–S08005.
- 2621 [31] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) p. S08003.
- 2623 [32] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 2626 [33] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 2629 [34] CERN, “The accelerator complex.” (2012).
- 2630 [35] G. C. Strong, *On the impact of selected modern deep-learning techniques to the performance and celerity of classification models in an experimental high-energy physics use case*, [2002.01427](https://arxiv.org/abs/2002.01427).
- 2633 [36] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022). <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idoveryview/>.
- 2635 [37] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, *JINST* **11** (2016) p. P04008 [[1512.01094](https://arxiv.org/abs/1512.01094)].
- 2637 [38] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021 (2019).
- 2639 [39] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design Report, Volume 1*, .

- 2641 [40] for the ATLAS Collaboration collaboration, *Status of the ATLAS Experiment*, Tech.
2642 Rep., Geneva (2010). 10.3204/DESY-PROC-2010-04/1.
- 2643 [41] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design
2644 Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 2645 [42] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design
2646 Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 2647 [43] ATLAS Collaboration, *The inner detector*, (2022).
2648 <https://atlas.cern/Discover/Detector/Inner-Detector>.
- 2649 [44] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*,
2650 ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 2651 [45] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,
2652 *JINST* **13** (2018) p. T05008 [[1803.00844](#)].
- 2653 [46] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor
2654 tracker*, *JINST* **9** (2014) p. P08009 [[1404.7473](#)].
- 2655 [47] ATLAS Collaboration, *Calorimeter*, (2022).
2656 <https://atlas.cern/Discover/Detector/Calorimeter>.
- 2657 [48] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, .
- 2658 [49] G. Artoni, *Muon momentum scale and resolution in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ in
2659 ATLAS*, Tech. Rep., Geneva (2016). 10.1016/j.nuclphysbps.2015.09.453.
- 2660 [50] ATLAS Collaboration, *Muon spectrometer*, (2022).
2661 <https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- 2662 [51] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur.
2663 Phys. J. C* **77** (2017) p. 317 [[1611.09661](#)].
- 2664 [52] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, (2021).
- 2665 [53] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts,
2666 Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 2667 [54] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense
2668 Environments*, ATL-PHYS-PUB-2015-006 (2015).
- 2669 [55] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in
2670 dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) p. 673 [[1704.07983](#)].

- 2671 [56] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon*
2672 *pixel detector*, *JINST* **9** (2014) p. P09009 [[1406.7690](#)].
- 2673 [57] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS*
2674 *Experiment*, .
- 2675 [58] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment*
2676 *in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332
2677 [[1611.10235](#)].
- 2678 [59] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for*
2679 *LHC Run 3*, ATL-PHYS-PUB-2019-015 (2019).
- 2680 [60] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency*
2681 *measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79**
2682 (2019) p. 970 [[1907.05120](#)].
- 2683 [61] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP*
2684 **04** (2008) p. 063 [[0802.1189](#)].
- 2685 [62] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur.Phys.J. C***72** (2012)
2686 p. 1896 [[1111.6097](#)].
- 2687 [63] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*
2688 *performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- 2689 [64] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*
2690 *the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 2691 [65] ATLAS Collaboration, *Jet energy scale measurements and their systematic*
2692 *uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
2693 *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].
- 2694 [66] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a*
2695 *New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008)
2696 p. 242001 [[0802.2470](#)].
- 2697 [67] ATLAS collaboration, *Measurement of the associated production of a Higgs boson*
2698 *decaying into b-quarks with a vector boson at high transverse momentum in pp*
2699 *collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
2700 p. 136204 [[2008.02508](#)].
- 2701 [68] ATLAS collaboration, *Performance of jet substructure techniques for large-R jets in*

- 2702 *proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **09** (2013)
2703 p. 076 [[1306.4945](#)].
- 2704 [69] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in*
2705 *13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019)
2706 p. 135 [[1807.09477](#)].
- 2707 [70] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* **06** (2009) p. 059
2708 [[0903.0392](#)].
- 2709 [71] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
2710 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010
2711 (2017).
- 2712 [72] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the*
2713 *Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012).
- 2714 [73] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*
2715 *using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024 (2016).
- 2716 [74] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in*
2717 *proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) p. 292
2718 [[1603.05598](#)].
- 2719 [75] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
2720 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys.*
2721 *J. C* **78** (2018) p. 903 [[1802.08168](#)].
- 2722 [76] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000)
2723 pp. 577–606 [[hep-ph/9912292](#)].
- 2724 [77] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**
2725 (2018) p. 030001.
- 2726 [78] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom*
2727 *and charm hadrons in the decays of top quarks and the fragmentation of high p_T jets*,
2728 ATL-PHYS-PUB-2014-008 (2014).
- 2729 [79] Nazar Bartosik, *Diagram showing the common principle of identification of jets*
2730 *initiated by b-hadron decays*, (2022).
2731 https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png.
- 2732 [80] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).

- 2733 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLLOTS/IDTR-2022-03/>.
- 2734 [81] S. Adorni Braccesi Chiassi, *Vector Boson Tagging in the Context of the Search for*
2735 *New Physics in the Fully Hadronic Final State*, Ph.D. thesis, Geneva U., 2021.
2736 10.13097/archive-ouverte/unige:158558.
- 2737 [82] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*
2738 *Process Interface for Event Generators*, ArXiv High Energy Physics - Phenomenology
2739 e-prints (2001) .
- 2740 [83] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*
2741 *standard format for Les Houches Event Files*, Computer Physics Communications
2742 **176** (2007) pp. 300–304.
- 2743 [84] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*
2744 *universal approximators*, Neural Networks **2** (1989) pp. 359–366.
- 2745 [85] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*
2746 *back-propagating errors*, nature **323** (1986) pp. 533–536.
- 2747 [86] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*
2748 *activity*, The bulletin of mathematical biophysics **5** (1943) pp. 115–133.
- 2749 [87] J. Hopfield, *Neural networks and physical systems with emergent collective*
2750 *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*
2751 *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 2752 [88] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*
2753 *networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 2754 [89] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv
2755 e-prints (2014) p. arXiv:1412.6980 [[1412.6980](https://arxiv.org/abs/1412.6980)].
- 2756 [90] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,
2757 Journal of High Energy Physics **2004** (2004) p. 040–040.
- 2758 [91] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*
2759 *carlo for heavy flavour hadroproduction*, Journal of High Energy Physics **2007** (2007)
2760 p. 126–126.
- 2761 [92] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*
2762 *shower simulations: the powheg method*, Journal of High Energy Physics **2007** (2007)
2763 p. 070–070.

- 2764 [93] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*
2765 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*
2766 *Physics* **2010** (2010) .
- 2767 [94] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015)
2768 p. 040 [[1410.8849](#)].
- 2769 [95] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
2770 ATL-PHYS-PUB-2016-020 (2016).
- 2771 [96] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An*
2772 *introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159
2773 [[1410.3012](#)].
- 2774 [97] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,
2775 ATL-PHYS-PUB-2014-021 (2014).
- 2776 [98] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013)
2777 p. 244 [[1207.1303](#)].
- 2778 [99] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*
2779 **462** (2001) p. 152.
- 2780 [100] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**
2781 (2010) p. 823 [[1005.4568](#)].
- 2782 [101] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.*
2783 *Instrum. Meth. A* **506** (2003) p. 250.
- 2784 [102] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS*
2785 *detector*, ATL-CONF-2014-018 (2014).
- 2786 [103] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter*
2787 *Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014 (2020).
- 2788 [104] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS*
2789 *Detector*, ATL-PHYS-PUB-2022-027 (2022).
- 2790 [105] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2791 *ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].
- 2792 [106] ATLAS Collaboration, *Observation of Higgs boson production in association with a*
2793 *top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173
2794 [[1806.00425](#)].

- 2795 [107] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*
2796 *using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *JHEP* **03**
2797 (2020) p. 145 [[1910.08447](#)].
- 2798 [108] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*
2799 *b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017).
- 2800 [109] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp*
2801 *collision dataset*, [2211.16345](#).
- 2802 [110] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,
2803 *Secondary vertex finding in jets with neural networks*, *The European Physical Journal*
2804 *C* **81** (2021) .
- 2805 [111] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,
2806 *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. arXiv:2002.08772
2807 [[2002.08772](#)].
- 2808 [112] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*
2809 *ATLAS detector*, ATL-PHYS-PUB-2017-011 (2017).
- 2810 [113] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
2811 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003 (2017).
- 2812 [114] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,
2813 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,
2814 *arXiv e-prints* (2018) p. arXiv:1806.01261 [[1806.01261](#)].
- 2815 [115] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,
2816 *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 2817 [116] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13\text{ TeV}$ data*,
2818 ATL-PHYS-PUB-2015-037 (2015).
- 2819 [117] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*
2820 *experiment using the 2015 and 2016 LHC proton–proton collision data at*
2821 *$\sqrt{s} = 13\text{ TeV}$* , *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 2822 [118] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*
2823 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, *arXiv e-prints* (2020)
2824 p. arXiv:2007.08294 [[2007.08294](#)].
- 2825 [119] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,

- 2826 *arXiv e-prints* (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 2827 [120] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2829 [121] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).
2830 <https://github.com/onnx/onnx>.
- 2831 [122] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].
- 2833 [123] 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) p. 136204 [[2008.02508](#)].
- 2837 [124] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039 [[1512.02572](#)].
- 2840 [125] 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) p. 076 [[1306.4945](#)].
- 2843 [126] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) p. 043 [[1002.2581](#)].
- 2846 [127] 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** (2014) p. 145 [[1406.3660](#)].
- 2849 [128] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al., *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* **72** (2012) p. 1889 [[1111.2034](#)].
- 2852 [129] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*, *JHEP* **10** (2012) p. 155 [[1206.3572](#)].
- 2854 [130] 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](#)].

- 2857 [131] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to*
2858 *associated WH and ZH production at hadron colliders*, *Phys. Rev. D* **68** (2003)
2859 p. 073003 [[hep-ph/0306234](#)].
- 2860 [132] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the*
2861 *Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) pp. 149–156
2862 [[hep-ph/0307206](#)].
- 2863 [133] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production*
2864 *at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.*
2865 **107** (2011) p. 152003 [[1107.1164](#)].
- 2866 [134] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects*
2867 *in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) p. 1868 [[1111.0761](#)].
- 2868 [135] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated*
2869 *WH production and decay at the LHC*, *JHEP* **04** (2014) p. 039 [[1312.1669](#)].
- 2870 [136] G. Ferrera, M. Grazzini and F. Tramontano, *Associated ZH production at hadron*
2871 *colliders: The fully differential NNLO QCD calculation*, *Phys. Lett. B* **740** (2015)
2872 pp. 51–55 [[1407.4747](#)].
- 2873 [137] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson*
2874 *at NNLO*, *JHEP* **06** (2016) p. 179 [[1601.00658](#)].
- 2875 [138] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,
2876 *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) p. 078
2877 [[1211.5015](#)].
- 2878 [139] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via*
2879 *gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) p. 065 [[1503.01656](#)].
- 2880 [140] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for*
2881 *gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) p. 082 [[1410.0217](#)].
- 2882 [141] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron*
2883 *Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) p. 023 [[1307.8122](#)].
- 2884 [142] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron*
2885 *colliders*, *Comput. Phys. Commun.* **184** (2013) pp. 998–1003 [[1210.5347](#)].
- 2886 [143] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte*
2887 *Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) p. 126 [[0707.3088](#)].

- 2888 [144] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair*
2889 *cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) p. 2930
2890 [[1112.5675](#)].
- 2891 [145] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with*
2892 *shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) p. 111
2893 [[0907.4076](#)].
- 2894 [146] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top*
2895 *quark production*, *Phys. Rev. D* **81** (2010) p. 054028 [[1001.5034](#)].
- 2896 [147] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for*
2897 *t-channel single top quark production*, *Phys. Rev. D* **83** (2011) p. 091503 [[1103.2792](#)].
- 2898 [148] E. Re, *Single-top Wt-channel production matched with parton showers using the*
2899 *POWHEG method*, *Eur. Phys. J. C* **71** (2011) p. 1547 [[1009.2450](#)].
- 2900 [149] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated*
2901 *production with a W^- or H^-* , *Phys. Rev. D* **82** (2010) p. 054018 [[1005.4451](#)].
- 2902 [150] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al.,
2903 *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) p. 007 [[0811.4622](#)].
- 2904 [151] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) p. 034
2905 [[1905.09127](#)].
- 2906 [152] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
2907 *Phys. Rev. Lett.* **108** (2012) p. 111601 [[1111.5206](#)].
- 2908 [153] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008)
2909 p. 039 [[0808.3674](#)].
- 2910 [154] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour*
2911 *dipole factorisation*, *JHEP* **03** (2008) p. 038 [[0709.1027](#)].
- 2912 [155] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton*
2913 *showers. The NLO case*, *JHEP* **04** (2013) p. 027 [[1207.5030](#)].
- 2914 [156] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson*
2915 *Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys.*
2916 *Rev. Lett.* **103** (2009) p. 082001 [[0903.2120](#)].
- 2917 [157] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43**
2918 (2016) p. 023001 [[1510.03865](#)].

- 2919 [158] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to*
2920 *Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* **03**
2921 (2012) p. 075 [[1112.5142](#)].
- 2922 [159] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo*
2923 *program for Higgs production in vector-boson fusion and Higgs strahlung at hadron*
2924 *colliders*, *Comput. Phys. Commun.* **195** (2015) pp. 161–171 [[1412.5390](#)].
- 2925 [160] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in*
2926 *Run-2*, ATL-PHYS-PUB-2015-022 (2015).
- 2927 [161] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS*
2928 *detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 2929 [162] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using*
2930 *ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2018-006
2931 (2018).
- 2932 [163] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using*
2933 *a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001 (2018).
- 2934 [164] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the*
2935 *ATLAS Detector*, ATL-PHYS-PUB-2014-013 (2014).
- 2936 [165] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark*
2937 *pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) p. 836
2938 [[1906.11005](#)].
- 2939 [166] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton*
2940 *distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) p. 204
2941 [[1412.3989](#)].
- 2942 [167] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton*
2943 *distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.*
2944 **D93** (2016) p. 033006 [[1506.07443](#)].
- 2945 [168] J. K. Anders and M. D’Onofrio, *V+Jets theoretical uncertainties estimation via a*
2946 *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 2947 [169] J. Pumplin et al., *New generation of parton distributions with uncertainties from*
2948 *global qcd analysis*, *JHEP* **07** (2002) p. 012 [[0201195](#)].
- 2949 [170] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The*

2950 *roostats project, arXiv preprint arXiv:1009.1003* (2010) .

2951 List of figures

<small>2952</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>2953</small>		
<small>2954</small>		
<small>2955</small>		
<small>2956</small>		
<small>2957</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>2958</small>		
<small>2959</small>		
<small>2960</small>		
<small>2961</small>		
<small>2962</small>		
<small>2963</small>	2.3 Higgs boson production cross sections as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [24]. 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>2964</small>		
<small>2965</small>		
<small>2966</small>		
<small>2967</small>	2.4 Higgs boson branching ratios as a function of Higgs mass (m_H) at $\sqrt{s} = 13$ TeV [24]. 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>2968</small>		
<small>2969</small>		
<small>2970</small>		
<small>2971</small>	3.1 An overview of the CERN accelerator complex [34]. 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>2972</small>		
<small>2973</small>		
<small>2974</small>		
<small>2975</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. [35].	30

2976	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. [36].	32
2982	3.4 Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [33]. A total of 139 fb^{-1} of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally.	34
2986	3.5 Average pile-up profiles measured by ATLAS during Run 2 [33]. Higher levels of pile-up are planned for Run 3.	34
2988	3.6 A 3D model of the entire ATLAS detector [40]. 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
2992	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 [43].	36
2994	3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [36].	37
2996	3.9 A schematic cross-sectional view of the ATLAS IBL [44].	38
2997	3.10 The ATLAS calorimeters [47]. 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
3000	3.11 The ATLAS muon spectrometer [50].	42
3001	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 [55].	46
3005	3.13 A sketch of electron reconstruction using the ATLAS detector [73]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.	50

3008	4.1 The truth b -hadron decay radius L_{xy} as a function of truth transverse 3009 momentum p_T for reconstructed b -jets in Z' events. Error bars show the 3010 standard deviation. The pixel layers are shown in dashed horizontal lines.	54
3011	4.2 Diagram of a typical b -jet (blue) which has been produced in an event 3012 alongside two light jets (grey) [79]. The b -hadron has travelled a significant 3013 distance (pink dashed line) from the primary interaction point (pink dot) 3014 before its decay. The large transverse impact parameter d_0 is a characteristic 3015 property of the trajectories of b -hadron decay products.	55
3016	4.3 At lower p_T (left) the decay length of the b -hadron is reduced, and the 3017 resulting decay tracks are less collimated. At higher p_T (right) the b -hadron 3018 decay length increases and the resulting decay tracks are more collimated and 3019 have less distance over which to diverge before reaching detector elements. 3020 As a result, the ID may be unable to resolve charged depositions from 3021 different particles, resulting merged clusters.	56
3022	4.4 b -hadron decay track reconstruction efficiency as a function of truth b -hadron 3023 p_T [80]. Nominal track reconstruction is shown in black, while the track 3024 reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver 3025 efficiency) is shown in green. For high- p_T b -hadrons, the ambiguity solver 3026 is overly aggressive in its removal of b -hadron decay tracks. Suggestions for 3027 the improvement of the track reconstruction efficiency in this regime by the 3028 loosening of cuts in the ambiguity solver are shown in blue and red.	57
3029	4.5 Hit multiplicities on the IBL (left) and the pixel layers (right) as a function 3030 of the p_T of the reconstructed track. Tracks from the weak decay of 3031 the b -hadron are shown in red, while fragmentation tracks (which are 3032 prompt) are in blue. Baseline tracks are those produced in the standard 3033 reconstruction described in Section 3.4.1, while pseudotracks represent the 3034 ideal performance of the ATLAS detector and are described in Section 3.4.1. 3035 Hit multiplicities on the pseudotracks decrease at high p_T due to the flight 3036 of the b -hadron before its decay. The baseline tracks have more hits than the 3037 pseudotracks, indicating that they are being incorrectly assigned additional 3038 hits on the inner layers of the detector.	59
3039	4.6 The fraction of hits which are shared on b -hadron decay tracks on the IBL 3040 (left), and the B-layer (right), as a function of the production radius of 3041 the b -hadron decay product. Pseudotracks represent the ideal performance 3042 given the ATLAS detector, see Section 3.4.1.	60

3043	4.7 The rate to assign good (left) and wrong (right) IBL hits as a function 3044 of b -hadron p_T for tracks using baseline tracking (black) and the modified 3045 version of the outlier removal procedure (red). For each track, the corre- 3046 sponding p_T bin is filled with the number of good or wrong hits and this 3047 value is averaged to show the overall rate.	62
3048	4.8 (left) b -hadron decay track d_0 pulls ($d_0/s(d_0)$) for baseline and modified 3049 GX2F tracks. (right) The absolute value of the d_0 pull as a function of 3050 b -hadron transverse momentum.	63
3051	5.1 A diagram displaying the logical flow of a single neuron with three inputs 3052 x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are 3053 summed. A bias term θ_0 is added, and the result z is passed to an activation 3054 function. Each neuron can be thought of as a logistic regression model. .	67
3055	5.2 The output of several common choices for the activation function $g(z)$ of 3056 an artificial neuron. The input z is the output of the dot product between 3057 the activation and the weights, plus a bias term.	68
3058	5.3 Rate of fake tracks as a function of jet transverse momentum (left) and 3059 $\Delta R(\text{track}, \text{jet})$ (right). The rate of fake tracks increases significantly as a 3060 function of p_T , and also increases as the distance to the jet axis decreases.	71
3061	5.4 The light-jet efficiency of the low level tagger SV1 as a function of b -jet 3062 efficiency for the nominal tracking setup (black) and for the case where 3063 fake tracks which are not from the decay of a b -hadron are removed. The 3064 light-jet efficiency is decreased, demonstrating that the presence of fake 3065 tracks is detrimental to the algorithm performance.	71
3066	5.5 (left) Normalised histogram of the model output separated for signal and 3067 fake tracks, and further separated by those tracks which are from the 3068 decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and 3069 tracks from the decay of a b -hadron (dashed line). The model is able to 3070 successfully discriminate between signal and fake tracks, and shows a very 3071 similar performance when looking specifically at tracks from the decay of a 3072 b -hadron.	76

3073	5.6 (left) Normalised histogram of the model output separated for tracks from 3074 the decay of a b -hadron and tracks from other sources. The two groups are 3075 further separated into those tracks which are fake. (right) The ROC curve 3076 for all tracks (solid line).	78
3077	5.7 The effect of applying the fake track identification algorithm alongside the 3078 b -hadron decay track identification on the jet tagging performance of SV1 3079 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and $400 \text{ GeV} < p_T < 1 \text{ TeV}$ 3080 (right). The nominal SV1 light-jet rejection (black) is compared to two 3081 working points of fake track removal, labelled ‘‘A’’ (red) and ‘‘B’’ (orange), 3082 which represent two different 2D working points of the track classification 3083 tools. Removal of fake tracks based on truth information is shown by the 3084 green curves.	79
3085	5.8 The effect of applying the fake track identification algorithm alongside 3086 the b -hadron decay track identification on the jet tagging performance of 3087 JetFitter for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with 3088 $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal JetFitter light-jet rejection 3089 (black) is compared to two working points of fake track removal, labelled ‘‘A’’ 3090 (red) and ‘‘B’’ (orange). Removal of fake tracks based on truth information 3091 is shown by the green curves.	80
3092	6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right). 3093 The existing approach utilises low-level algorithms (shown in blue), the 3094 outputs of which are fed into a high-level algorithm (DL1r). Instead 3095 of being used to guide the design of the manually optimised algorithms, 3096 additional truth information from the simulation is now being used as 3097 auxiliary training targets for GN1. The solid lines represent reconstructed 3098 information, whereas the dashed lines represent truth information.	84
3099	6.2 The flow of information through a graph neural network layer for (left) a full 3100 implementation of the layer and (right) a deep sets model [115]. Reproduced 3101 from Ref. [114].	87
3102	6.3 The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of 3103 n_{tracks} , where each track is described by 21 track features ($n_{\text{tf}} = 21$). The 3104 jet features are copied for each of the tracks, and the combined jet-track 3105 vectors of length 23 form the inputs of GN1.	92

3106	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.	93
3107		
3108	These representations are then used to populate the node features of a	
3109	fully connected graph network. After the graph network, the resulting node	
3110	representations are used to predict the jet flavour, the track origins, and	
3111	the track-pair vertex compatibility.	93
3112	6.5 Comparison between the DL1r and GN1 b -tagging discriminant D_b for	
3113	jets in the $t\bar{t}$ sample. The 85% WP and the 60% WP are marked by the	
3114	solid (dashed) lines for GN1 (DL1r), representing respectively the loosest	
3115	and tightest WPs used by analyses. A value of $f_c = 0.018$ is used in the	
3116	calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. The distributions	
3117	of the different jet flavours have been normalised to unity area.	98
3118	6.6 The c -jet (left) and light-jet (right) rejections as a function of the b -jet	
3119	tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV. The	
3120	ratio with respect to the performance of the DL1r algorithm is shown in	
3121	the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b	
3122	for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. Binomial error bands	
3123	are denoted by the shaded regions. At b -jet tagging efficiencies less than	
3124	$\sim 75\%$, the light-jet rejection becomes so large that the effect of the low	
3125	number of jets is visible. The lower x -axis range is chosen to display the	
3126	b -jet tagging efficiencies usually probed in these regions of phase space. .	100
3127	6.7 The c -jet (left) and light-jet (right) rejections as a function of the b -jet	
3128	tagging efficiency for jets in the Z' sample with $250 < p_T < 5000$ GeV. The	
3129	ratio with respect to the performance of the DL1r algorithm is shown in	
3130	the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b	
3131	for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. Binomial error bands	
3132	are denoted by the shaded regions. At b -jet tagging efficiencies less than	
3133	$\sim 20\%$, the light-jet rejection becomes so large that the effect of the low	
3134	number of jets is visible. The lower x -axis range is chosen to display the	
3135	b -jet tagging efficiencies usually probed in these regions of phase space. .	101

3136	6.8 The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. GN1 demonstrates improved performance with respect to DL1r across the p_T range shown. Binomial error bands are denoted by the shaded regions.	102
3142	6.9 The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. At c -jet tagging efficiencies than $\sim 25\%$, the light-jet rejection becomes so large that the effect of the low number of jets is visible. The lower x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.	104
3149	6.10 The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The lower x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.	104
3155	6.11 A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). 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.	106
3162	6.12 A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities p_b , p_c and p_u are rounded to three significant figures.	106

3166	6.13 The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 65\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.	108
3177	6.14 The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. At b -jet tagging efficiencies less than $\sim 25\%$, the light-jet rejection become so large that the effect of the low number of jets are visible. The lower x -axis range is chosen to display the b -jet tagging efficiencies usually probed in these regions.	109
3188	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. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.	112
3193	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 . 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.	113

3198	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.	115
3205	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.	117
3211	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.	119
3217	7.1 Diagrams of the signal process (top) and the 0-lepton and 1-lepton $t\bar{t}$ back- grounds (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.	129
3221	7.2 Normalised leading large- R jet mass distribution from Z and $W+hf$ pro- cesses in the HP SR of the 0-lepton channel. 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.	139
3225	7.3 The comparison on m_J shapes between SHERPA and MADGRAPH samples from $V+hf$ process in p_T^V inclusive signal regions. The Kolmogorov-Smirnov test and χ^2/ndf are shown on the plots.	140
3228	7.4 Normalised leading large- R jet mass distribution from WZ and ZZ process, merged among all the signal regions and lepton channels. The renormalisa- tion 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.	142

3232	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. 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.	143
3237	7.6	The m_J post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$. The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions are shown as filled histograms. The Higgs boson signal ($m_H = 125 \text{ GeV}$) is shown as a filled histogram and is normalised to the signal yield extracted from data ($\mu_{VH}^{bb} = 0.72$), and as an unstacked unfilled histogram, scaled by the SM prediction times a factor of two. The size of the combined on the sum of the fitted signal and background is shown in the hatched band. The highest bin contains the overflow.	149
3247	7.7	The m_J post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$. The background contributions after the likelihood fit are shown as filled histograms. The Higgs boson signal ($m_h = 125 \text{ GeV}$) is shown as a filled histogram on top of the fitted backgrounds normalised to the signal yield extracted from data ($\mu_{VH}^{bb} = 0.72$), and unstacked as an unfilled histogram, scaled by the SM prediction times a factor of 2. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The highest bin in the distributions contains the overflow.	150

3258 7.8 m_J distribution in data after subtraction of all backgrounds except for the
 3259 WZ and ZZ diboson processes. The contributions from all lepton channels
 3260 and signal regions are summed and weighted by their respective values of
 3261 the ratio of fitted Higgs boson signal and background yields. The expected
 3262 contribution of the associated WH and ZH production of a SM Higgs
 3263 boson with $m_H = 125$ GeV is shown scaled by the measured combined
 3264 signal strength ($\mu_{VH}^{bb} = 0.72$). The diboson contribution is normalised to
 3265 its best-fit value of $\mu_{VZ}^{bb} = 0.91$. The size of the combined statistical and
 3266 systematic uncertainty is indicated by the hatched band. This error band is
 3267 computed from a full signal-plus-background fit including all the systematic
 3268 uncertainties defined in Section 7.2, except for the VH/VZ experimental
 3269 and theory uncertainties.

3270 7.9 Signal strength compatibility test between the (3+1)-POI fit (with the three
 3271 lepton channels splitted) and the default (1+1)-POI fit. The compatibility
 3272 of the three channels is evaluated via a χ^2 difference test and results in a
 3273 p-value of 49%.

3274 7.10 Impact of systematic uncertainties on the fitted VH signal-strength pa-
 3275 rameter μ_{VH}^{bb} sorted in decreasing order. The boxes show the variations
 3276 of $\hat{\mu}$, referring to the top x -axis, when fixing the corresponding individual
 3277 nuisance parameter to its post-fit value modified upwards or downwards by
 3278 its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$, and repeating the fit. The impact of
 3279 up- and down-variations can be distinguished via the dashed and plane box
 3280 fillings. The yellow boxes show the pre-fit impact (top x -axis) by varying
 3281 each nuisance parameter by ± 1 . The filled circles show the deviation of the
 3282 fitted value for each nuisance parameter, $\hat{\theta}$, from their nominal input value
 3283 θ_0 expressed in standard deviations with respect to their nominal uncer-
 3284 tainties $\Delta\theta$ (bottom x -axis). The error bars show the post-fit uncertainties
 3285 on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show
 3286 the fitted values and uncertainties of the normalization parameters that are
 3287 freely floating in the fit. Pre-fit, these parameters have a value of one.

3288 List of tables

<small>3289</small>	2.1 The fermions of the SM [14]. 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>3290</small>	<small>3291</small>	14				
<small>3292</small>	2.2 The bosons of the SM [14]. 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>3293</small>	<small>3294</small>	<small>3295</small>	15			
<small>3296</small>	3.1 Overview of the different LHC runs [32, 33]. The average number of interac- tions per bunch-crossing is denoted as $\langle \mu \rangle$ (see Section 3.2.3), and is here averaged over the entire run. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.	<small>3297</small>	<small>3298</small>	<small>3299</small>	28			
<small>3300</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 [55]. 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>3301</small>	<small>3302</small>	<small>3303</small>	<small>3304</small>	<small>3305</small>	<small>3306</small>	70
<small>3307</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 [55], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.	<small>3308</small>	<small>3309</small>	<small>3310</small>	<small>3311</small>	<small>3312</small>	74	

3313	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 [55]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.	75
3319	5.4 Hyperparameter for the track classification model	75
3320	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.	77
3324	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).	78
3328	6.1 Input features to the GN1 model. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [55], 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.	90
3335	6.2 A summary of GN1’s different classification networks used for the different training objectives. The hidden layers column contains a list specifying the number of neurons in each layer.	94
3338	6.3 The area under the ROC curves (AUC) for the track classification from GN1, compared to a standard multilayer perceptron (MLP) trained on a per-track basis. The unweighted mean AUC over the origin classes and weighted mean AUC (using as a weight the fraction of tracks from the given origin) is provided. GN1, which uses an architecture that allows track origins to be classified in a conditional manner as discussed in Section 6.4.3, outperforms the MLP model for both $t\bar{t}$ and Z' jets.	115

3345	7.1 Summary of the definition of the analysis regions. Signal enriched regions 3346 are marked with the label SR. There are regions with relatively large signal 3347 purity (HP SR) and with low purity (LP SR). Background enriched regions 3348 are marked with the label CR. The shorthand ‘‘add’’ stands for additional 3349 small- R jets, i.e. number of small- R jets not matched to the Higgs-jet 3350 candidate.	123
3351	7.2 Signal and background processes with the corresponding generators used for the 3352 nominal samples. If not specified, the order of the cross-section calculation refers to 3353 the expansion in the strong coupling constant (α_s). (★) The events were generated 3354 using the first PDF in the NNPDF3.0NLO set and subsequently reweighted to the 3355 PDF4LHC15NLO set [157] using the internal algorithm in POWHEG-Box v2. (†) 3356 The NNLO(QCD)+NLO(EW) cross-section calculation for the $pp \rightarrow ZH$ process 3357 already includes the $gg \rightarrow ZH$ contribution. The $qq \rightarrow ZH$ process is normalised 3358 using the cross-section for the $pp \rightarrow ZH$ process, after subtracting the $gg \rightarrow ZH$ 3359 contribution. An additional scale factor is applied to the $qq \rightarrow VH$ processes 3360 as a function of the transverse momentum of the vector boson, to account for 3361 electroweak (EW) corrections at NLO. This makes use of the VH differential 3362 cross-section computed with HAWK [158, 159].	125
3363	7.3 Event selection requirements for the boosted VH , $H \rightarrow b\bar{b}$ analysis channels 3364 and sub-channels. Various channel dependent selections are used to 3365 maximise the signal efficiency and reduce the number of background events 3366 in each analysis region.	127
3367	7.4 Different sources of uncertainty (i.e. variations in the model) considered for 3368 the $V+jets$ background, and the corresponding implementation. For each 3369 uncertainty, acceptance and shape uncertainties are derived.	133
3370	7.5 0-lepton $W+jets$ nominal sample flavour composition and total event yield. . .	135
3371	7.6 1-lepton $W+jets$ nominal sample flavour composition and total event yield. . .	135
3372	7.7 0-lepton $Z+jets$ nominal sample flavour composition and total event yield. . .	135
3373	7.8 2-lepton $Z+jets$ nominal sample flavour composition and total event yield. . .	135

3374	7.9 V+jets acceptance uncertainties. $W+$ jets SR and CR uncertainties marked 3375 with a superscript \dagger are correlated. The 1L $W+$ jets H/M uncertainty 3376 marked by * is applied as independent and uncorrelated NPs in both HP 3377 and LP signal regions. The 0L $W+$ jets Wbb Norm uncertainty is only 3378 applied when a floating normalisation for Wbb cannot be obtained from 3379 the 1L channel. A 30% uncertainty for $Z \rightarrow b\bar{b}$ norm is applied in the 1L 3380 channel when a floating normalisation for $Z \rightarrow b\bar{b}$ cannot be obtained from 3381 the 0L or 2L channels.	137
3382	7.10 Diboson acceptance uncertainties. All uncertainties except channel extrap- 3383 olation uncertainties are fully correlated between ZZ and WZ processes 3384 and channels.	142
3385	7.11 Factors applied to the nominal normalisations of the $t\bar{t}$, $W+hf$, and $Z+hf$ 3386 backgrounds, as obtained from the likelihood fit. The errors represent the 3387 combined statistical and systematic uncertainties.	147
3388	7.12 Breakdown of the observed absolute contributions to the uncertainty on the 3389 signal strength μ_{VH}^{bb} obtained from the (1+1)-POI fit. The average impact 3390 represents the average between the positive and negative uncertainties on 3391 μ_{VH}^{bb} . The sum in quadrature of the systematic uncertainties attached to the 3392 categories differs from the total systematic uncertainty due to correlations.	156