

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**

September 13, 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

no more than 300 words

This thesis presents work on improvement of the tracking and b -tagging algorithms used by ATLAS to perform b -tagging at high transverse momenta. Novel approaches, such as the classification of track origins and the application of graph neural networks have been successfully employed to markedly improve the b -tagging performance at high p_T . Work on analysis of $H \rightarrow b\bar{b}$ decays was also performed using 13 TeV proton-proton collision data from the course of Run 2.

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.

Perhaps you might want to point out that Peter Higgs provided help

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

Brian Moser provided insight on the fit.

Contents

1	Theoretical Framework	9
1.1	The Standard Model	10
1.1.1	Quantum Electrodynamics	11
1.1.2	Quantum Chromodynamics	12
1.1.3	The Electroweak Sector	13
1.2	The Higgs Mechanism	14
1.2.1	Electroweak Symmetry Breaking	15
1.2.2	Fermionic Yukawa Coupling	18
1.2.3	Higgs Phenomenology	20
2	The Large Hadron Collider and the ATLAS Detector	23
2.1	Coordinate System & Collider Definitions	24
2.1.1	ATLAS Coordinate System	24
2.1.2	Track Parameterisation	25
2.1.3	Hadron Collider Definitions	27
2.2	The ATLAS Detector	29
2.2.1	Inner Detector	30
2.2.2	Calorimeters	33
2.2.3	Muon Spectrometer	35
2.2.4	The Trigger	35
2.3	Reconstructed Physics Objects	36
2.3.1	Tracks	37
2.3.2	Vertices	39
2.3.3	Jets	40
2.3.4	Leptons	41
2.3.5	Missing Transverse Momentum	43

3 Tracking and b-tagging	44
3.1 b -hadron Reconstruction	45
3.1.1 Decay Topology	45
3.1.2 Challenges Facing b -hadron Reconstruction	47
3.2 Investigations into High p_T b -hadron Tracking	50
3.2.1 Pseudotracking	51
3.2.2 Global χ^2 Fitter Outlier Removal	52
3.3 Conclusion	55
4 Track Classification MVA	56
4.1 Machine Learning Background	56
4.1.1 Neural Networks	58
4.1.2 Training with Gradient Descent	59
4.2 Track Truth Origin Labelling	60
4.3 Fake Track Identification Tool	61
4.3.1 Datasets	62
4.3.2 Model Inputs	64
4.3.3 Results	64
4.4 Conclusion	67
5 Graph Neural Network Flavour Tagger	69
5.1 Motivation	70
5.2 Graph Neural Network Theory	73
5.3 Experiemental Setup	73
5.3.1 Datasets	73
5.4 Model Architecture	73
5.4.1 Model Inputs	73
5.4.2 Auxiliary Training Objectives	74
5.4.3 Architecture	76
5.4.4 Training	80
5.5 Results	81
5.5.1 b -tagging Performance	82
5.5.2 c -tagging Performance	86
5.5.3 Jet Display Diagrams	88
5.5.4 Ablations	88
5.5.5 Inclusion of Low-Level Vertexing Algorithms	93

5.5.6	Vertexing Performance	93
5.5.7	Track Classification Performance	97
5.6	Extensions	99
5.6.1	Looser Track Selection	99
5.7	Conclusion	99
6	Boosted VHbb Analysis	101
6.1	Analysis Overview	102
6.1.1	Simulated Samples	103
6.1.2	Selection Criteria	103
6.1.3	Signal and Control Regions	105
6.2	Modelling & Systematic Uncertainties	105
6.2.1	Overview	105
6.2.2	Sources of Systematic Uncertainties	105
6.2.3	Shape Uncertainties	107
6.2.4	Acceptance Uncertainties	107
6.2.5	Vector Boson + Jets Modelling	109
6.2.6	Diboson Modelling	111
6.3	Fit Model	111
6.4	Results	111
6.4.1	Pre-fit Yields	111
6.4.2	Post-fit Results	111
6.5	Improved b -tagging using GNNs	112
6.6	Conclusion	112
7	Conclusion	114
Bibliography		115
List of figures		124
List of tables		132

² Chapter 1

³ Theoretical Framework

⁴ The Standard Model (SM) of particle physics is the theory describing all known
⁵ elementary particles and their interactions via three of the four fundamental forces.
⁶ Developed by merging the successful theories of quantum mechanics and relativity
⁷ in the second half of the 20th century, the SM's position today at the centre of our
⁸ understanding of the nature of the universe is firmly established by an unparalleled
⁹ level of agreement between the predictions from the model and experimental results
¹⁰ [2, 3].

¹¹ The SM has predicted the discovery of the top and bottom quarks [4–6], the W
¹² and Z bosons [7], and the tau neutrino [8]. The last missing piece of the SM to be
¹³ discovered was the Higgs boson, first theorised in the 1960s [9–11], and eventually
¹⁴ observed at the LHC in 2012 [12, 13]. After its discovery, much ongoing work has
¹⁵ been carried out performing detailed measurements of its mass and interactions with
¹⁶ other particles.

¹⁷ This thesis describes various efforts in improving the understanding of the Higgs
¹⁸ boson and its coupling to b - and c -quarks, primarily through the improvement of
¹⁹ algorithmic tools, discussed in Chapters 4 and 5. Analysis work on $H \rightarrow b\bar{b}$
²⁰ decays is also described in Chapter 6. In this chapter, an overview of the SM is
²¹ given in Section 1.1, and a more detailed discussion of the Higgs sector and Higgs
²² phenomenology is provided in Section 1.2.

²³ 1.1 The Standard Model

²⁴ The SM is formulated in the language of Quantum Field Theory (QFT). In this
²⁵ framework, particles are localised excitations of corresponding quantum fields, which
²⁶ are operator-valued distributions across spacetime.

²⁷ Central to QFT is the Lagrangian density which describes the kinematics and
²⁸ dynamics of a field. Observations of conserved quantities are linked, via Noether's
²⁹ theorem, to symmetries which are expressed by the Lagrangian. Alongside Global
³⁰ Poincaré symmetry, the SM Lagrangian observes a local non-Abelian $SU(3)_C \otimes$
³¹ $SU(2)_L \otimes U(1)_Y$ gauge symmetry. Gauge symmetries leave observable properties of
³² the system unchanged when the corresponding gauge transformations are applied
³³ to the fields. The full Lagrangian of the SM can be broken up into distinct terms
³⁴ corresponding to the different sectors, as in Eq. (1.1). An overview of each sector is
³⁵ given in the following chapters.

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} \quad (1.1)$$

³⁶ The SM provides a mathematical description of how the four fundamental forces
³⁷ interact with the matter content of the universe. The SM contains 12 spin-1/2
fermions, listed in Table 1.1, and 5 bosons listed in Table 1.2.

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 1.1: The half-integer spin 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.

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

Table 1.2: The integer spin 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 particle is the only scalar boson in the SM.

³⁹ 1.1.1 Quantum Electrodynamics

⁴⁰ Quantum electrodynamics (QED) is the relativistic quantum theory which describes
⁴¹ the interaction between lighted and charged matter. Consider a Dirac spinor field
⁴² $\psi = \psi(x)$ and its adjoint $\bar{\psi} = \psi^\dagger \gamma^0$, where ψ^\dagger denotes the Hermitian conjugate of ψ .
⁴³ The field ψ describes fermionic spin- $1/2$ particle, for example an electron. The Dirac
⁴⁴ Lagrangian density is

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\cancel{\partial} - m)\psi, \quad (1.2)$$

⁴⁵ where $\cancel{\partial} = \gamma^\mu \partial_\mu$ denotes the contraction with the Dirac gamma matrices γ^μ (summa-
⁴⁶ tion over up-down pairs of indices is assumed). Application of the Euler-Lagrange
⁴⁷ equation on Eq. (1.2) yields the Dirac equation

$$(i\cancel{\partial} - m)\psi = 0. \quad (1.3)$$

⁴⁸ Suppose some fundamental symmetry that requires invariance under a local $U(1)$
⁴⁹ gauge transformation

$$\psi \rightarrow \psi' = \psi e^{-iq\alpha(x)}, \quad (1.4)$$

⁵⁰ where α varies over every spacetime point x . Under this transformation, the Dirac
⁵¹ equation transforms as

$$(i\cancel{\partial} - m)\psi e^{-iq\alpha(x)} + q\cancel{\partial}\alpha(x)\psi e^{-iq\alpha(x)} = 0. \quad (1.5)$$

52 For the Dirac equation to remain invariant under the transformation in Eq. (1.4),
 53 a new field A_μ which transforms as $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x)$ must be added. The
 54 transformed interaction term

$$-qA\psi \rightarrow -qA\psi e^{-iq\alpha(x)} - q\partial\alpha(x)\psi e^{-iq\alpha(x)} \quad (1.6)$$

55 will then cancel the asymmetric term in Eq. (1.5) as required. The $U(1)$ invariant
 56 Lagrangain can therefore be constructed by adding an interaction between ψ and
 57 A_μ to Eq. (1.2). For completeness, the kinetic term for the the new field A_μ is
 58 also added in terms of $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, which is trivially invariant under the
 59 transformation in Eq. (1.4). The interaction term is typically absorbed into the
 60 covariant derivative $D_\mu = \partial_\mu + iqA_\mu$, thus named as it transforms in the same way as
 61 the field ψ . Collecting these modifications to Eq. (1.2) yields the QED Lagrangain

$$\mathcal{L}_{\text{QED}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(iD^\mu - m)\psi. \quad (1.7)$$

62 A quadratic term $A_\mu A^\mu$ is not invariant and therefore the the field A_μ must be
 63 massless. Requiring invariance under local $U(1)$ gauge transformations necessitated
 64 the addition of a new field A_μ , interpreted as the photon field, which interacts with
 65 charged matter. In the SM, the QED Lagrangian is absorbed into the electroweak
 66 sector, discussed in Section 1.1.3.

67 1.1.2 Quantum Chromodynamics

68 Quantum Chromodynamics (QCD) is the study of quarks, gluons and their interac-
 69 tions. Quarks and gluons carry colour charge, which comes in three kinds, called
 70 red, green and blue. While the $U(1)$ symmetry group in Section 1.1.1 was Abelian,
 71 the QCD Lagrangian is specified by requiring invariance under transformations from
 72 the non-Abelian $SU(3)$ group, making it a Yang-Mills theory [15] which requires the
 73 addition of self-interacting gauge fields. The infinitesimal $SU(3)$ group generators
 74 are given by $T_a = \lambda_a/2$, where λ_a are the eight Gell-Mann matrices. These span the
 75 space of infinitesimal group transformations and do not commute with each other,
 76 instead satisfying the commutation relation

$$[T_a, T_b] = if_{abc}T_c, \quad (1.8)$$

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

$$G_{\mu\nu}^a = \partial_\mu A_\nu - \partial_\nu A_\mu - g_s f^{abc} A_\mu^b A_\nu^c, \quad (1.9)$$

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

$$D_\mu = \partial_\mu + i g_s T_a A_\mu^a. \quad (1.10)$$

84 The full QCD Lagrangian is then given by

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} + q_k (i \not{D} - m_k) q_k. \quad (1.11)$$

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

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

95 1.1.3 The Electroweak Sector

96 The weak and electromagnetic forces are unified in the Glashow-Weinberg-Salam
 97 (GWS) model of electroweak interaction [16–18]. The Lagrangian is specified by
 98 requiring invariance under the symmetry group $SU(2)_L \otimes U(1)_Y$, as motivated by a
 99 large amount of experimental data. Here, $SU(2)_L$ is referred to as weak isospin and
 100 $U(1)_Y$ as weak hypercharge.

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

$$[T_a, T_b] = i\varepsilon_{abc}T_c. \quad (1.12)$$

- 103 The generator of $U(1)_Y$ is $Y = 1/2$. Each generator corresponds to a gauge field,
104 which, after symmetry breaking (discussed in Section 1.2), give rise to the massive
105 vector bosons, W^\pm and Z , and the massless photon. The massive vector bosons are
106 the carriers of the weak force, and are unique to the weak sector. Due to the mass
107 of the force carriers, the weak force has a short range and so it appears weak even
108 though its intrinsic strength is comparable to that of QED.

- 109 The weak force violates parity conservation [19–21], i.e. invariance under parity
110 transformations (mirror reflections). Only left handed fermions participate in the
111 weak interaction. Since there is no other force through which neutrinos interact
112 with other particles, there are no right handed neutrinos in the standard model.
113 Furthermore, the weak sector exhibits CP violation. CP violation is one of the three
114 necessary Sakharov conditions required to produce baryon asymmetry in the universe.
115 Since the SM alone does not appear to have enough CP violation to generate the
116 cosmologically observed matter-antimatter asymmetry, looking for signs of more
117 experimental CP violation is considered to be a promising way to discover new
118 physics.

- 119 The charge operator Q can be written

put somewhere else?

$$Q = T_3 + Y. \quad (1.13)$$

120 1.2 The Higgs Mechanism

- 121 The Brout-Englert-Higgs mechanism (henceforth just “Higgs mechanism”) is the
122 mechanism through which the fundamental particles of the SM acquire mass [9–11].
123 Experimentally it was known that the weak force had a weak effective strength,
124 which was suggestive of a massive mediating gauge particle. However, directly
125 adding mass to the weak gauge bosons violates the non-Abelian symmetry of the
126 SM. Instead, the gauge bosons can gain mass through the interaction with a scalar

127 Higgs field which results from the spontaneous breakdown of symmetry as discussed
128 in Section 1.2.1. Similarly, the Higgs mechanism gives mass to the fermions, as
129 discussed in Section 1.2.2. Section 1.2.3 described some basic phenomenology of the
130 Higgs particle relevant to hadron colliders.

131 **1.2.1 Electroweak Symmetry Breaking**

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

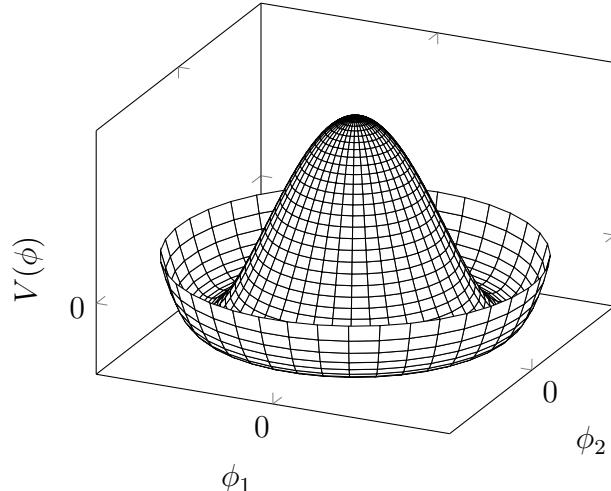


Figure 1.1: The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$, with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. The SM Higgs mechanism relies on a complex scalar doublet with a corresponding 5-dimensional potential.

138 Consider gauge fields from the local $SU(2)_L \otimes U(1)_Y$ symmetry group discussed in
139 Section 1.1.3 coupled to a complex scalar field $\phi = \phi(x)$. The scalar field ϕ transforms
140 as a weak isospin doublet. Omitting the kinetic term of the gauge fields, and writing
141 $\phi^2 \equiv \phi^\dagger \phi$, the Lagrangain is

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - [\mu^2 \phi^2 + \lambda \phi^4], \quad (1.14)$$

¹⁴² where the covariant derivative is given by

$$D_\mu = \partial_\mu + igA_\mu^a T^a + ig'B_\mu, \quad (1.15)$$

¹⁴³ and T^a are the generators of $SU(2)$. The potential term $V(\phi)$ is made up of a
¹⁴⁴ quadratic and quartic term in the scalar field ϕ , which each contain an arbitrary
¹⁴⁵ parameter, respectively λ and μ . The quartic term gives the field self-interaction, and
¹⁴⁶ cannot be negative as this would lead to a potential that was unbounded from below.
¹⁴⁷ The quadratic term can be positive or negative. In the case where the quadratic
¹⁴⁸ term is positive, it is interpreted as a mass term for the scalar field. By choosing
¹⁴⁹ $\mu^2 < 0$ the field becomes unphysical due to its negative mass. In order to obtain a
¹⁵⁰ physical interpretation of the Lagrangian in Eq. (1.14) for the case where $\mu^2 < 0$, the
¹⁵¹ field ϕ is expanded around the vacuum state. The vacuum expectation value (VEV)
¹⁵² is expected value of the field ϕ which minimises the potential $V(\phi)$ (equivalently
¹⁵³ the expected value of the field operator ϕ when the system is in a vacuum state,
¹⁵⁴ $|\langle\phi\rangle_0|^2 \equiv |\langle 0|\phi|0\rangle|^2 \equiv \phi_0^2$). Minimising the potential gives a VEV of

$$\phi_0^2 = -\mu^2/\lambda = v^2. \quad (1.16)$$

¹⁵⁵ Due to the shape of the potential in Fig. 1.1, there is degeneracy in the direction
¹⁵⁶ that the complex doublet ϕ points. As all the different vacuum states minimise
¹⁵⁷ the potential and therefore yield identical physics, one can arbitrarily choose the
¹⁵⁸ state to lie along the second component of the doublet. Application of Eq. (1.13)
¹⁵⁹ shows this choice is manifestly invariant under the charge operator. This allows
¹⁶⁰ the identification of the unbroken subgroup $U(1)_Q$, under which the ground state is
¹⁶¹ invariant. The generator of $U(1)_Q$ is the charge operator Q .

¹⁶² Adding the particle content back to the theory by expanding the field around
¹⁶³ the vacuum state, and making a transformation to the unitary gauge to remove
¹⁶⁴ unphysical would-be Nambu-Goldstone modes (which arise in the context of global
¹⁶⁵ symmetries [22, 23]), yields

$$\phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v + H \end{bmatrix}, \quad (1.17)$$

¹⁶⁶ Where H is a real scalar field, the *true vacuum* Higgs field. Substituting this
¹⁶⁷ into Eq. (1.14) and identifying physical fields from the quadratic terms of linear

168 combinations of unphysical fields, one can write the physical fields W_μ^\pm , Z_μ and A_μ
169 in terms of the original fields A_μ^a and B_μ . This gives

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2) \quad \begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix} = \begin{bmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{bmatrix} \begin{bmatrix} B_\mu \\ A_\mu^3 \end{bmatrix}. \quad (1.18)$$

170 where θ_W is the weak mixing angle defined by

$$\cos \theta_W = \frac{g}{\sqrt{g^2 + g'^2}}. \quad (1.19)$$

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

$$m_W = \frac{1}{2}gv \quad m_Z = \frac{m_W}{\cos \theta_W}, \quad (1.20)$$

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

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

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

190 Spontaneous symmetry breaking has modified the original symmetry group of the SM
191 $SU(2)_L \times U(1)_Y \rightarrow SU(3)_C \otimes U(1)_Q$. Three broken generators from the symmetry

192 group $SU(2)_L \times U(1)_Y$ have been absorbed into the definition of the physical weak
193 vector bosons, giving them mass. The same methodology can be used to generate
194 the fermion masses, as shown in the next section.

195 1.2.2 Fermionic Yukawa Coupling

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

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

205 Lepton Masses

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

$$\mathcal{L}_{\text{Yukawa}} = -g_f(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L). \quad (1.21)$$

208 where $\psi_L = (\nu_L, e_L)$ and $\psi_R = e_R$ for the first generation leptons. After spontaneous
209 symmetry breaking (see Section 1.2.1), the scalar Higgs field in unitary gauge
210 Eq. (1.17) consists of a VEV and the true vacuum Higgs field H . Substituting this
211 in to Eq. (1.21) yields

$$\mathcal{L}_{\text{Yukawa}} = -\frac{v g_e}{\sqrt{2}} \bar{e} e - \frac{g_e}{\sqrt{2}} \bar{e} H e, \quad (1.22)$$

212 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
213 the first term in Eq. (1.22) which is quadratic in the electron field, and can therefore
214 be identified as the electron mass term. An interaction term between the electron

215 field e and the true vacuum Higgs field H is also present. Mass is generated for the
216 other lepton generations in the same way.

217 **Quark Masses**

218 The down-type quarks acquire their mass analogous to the leptons, with $\psi_L = (u_L, d_L)$
219 and $\psi_R = d_R$ for the first quark generation. Mass is generated for the up-type quarks
220 using the conjugate field to ϕ which transforms under $SU(2)$ as a doublet with
221 $Y = -1/2$. The conjugate field $\tilde{\phi}$ is constructed as

$$\tilde{\phi} = i\sigma_2 \phi^\dagger = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1^\dagger \\ \phi_2^\dagger \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} v + H \\ 0 \end{bmatrix}, \quad (1.23)$$

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

$$\mathcal{L}_{\text{Yukawa}}^{\text{up}} = -g_q (\bar{\psi}_L \tilde{\phi} \psi_R + \bar{\psi}_R \tilde{\phi}^\dagger \psi_L) \quad (1.24)$$

225 Considering the first generation of up-type quarks with $\psi_L = (u_L, d_L)$ and $\psi_R = u_R$,
226 substitution into Eq. (1.24) yields

$$\mathcal{L}_{\text{Yukawa}}^{\text{up}} = -\frac{vg_u}{\sqrt{2}} \bar{u}u - \frac{g_u}{\sqrt{2}} \bar{u}Hu. \quad (1.25)$$

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

$$\begin{bmatrix} \tilde{d} \\ \tilde{s} \\ \tilde{b} \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}, \quad (1.26)$$

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

237 1.2.3 Higgs Phenomenology

238 As previous discussed in this section, the Higgs plays a key role in the SM, giving
239 mass to fundamental particles. The Higgs itself gains mass through self interaction.
240 The strength of the coupling between the Higgs and some other particle is propor-
241 tional to that particle's mass. This fact dictates which production mechanisms and
242 decay modes are dominant at the LHC. The cross sections for different production
243 mechanisms at a centre of mass energy $\sqrt{s} = 13 \text{ TeV}$ are shown as a function of the
244 Higgs mass m_H in Fig. 1.2. Production occurs mainly through gluon-gluon fusion
245 ($pp \rightarrow H$), mediated by a virtual top quark loop. Vector boson fusion ($pp \rightarrow qqH$) is
246 the second most dominant production mechanism, in which a pair of W or Z bosons
247 fuse to produce a Higgs after being radiated by two quarks, which also occur in the
248 final state. Next most common is the associated production of a Higgs and a vector
249 boson ($pp \rightarrow VH$), in which a pair of quarks fuse to produce a single W or Z boson
250 which radiates a Higgs. Although gluon-gluon fusion is the dominant production
251 mode, the associated Higgs production with a vector boson has the advantage of
252 leading to a more conspicuous final state due to the likelihood of the vector bosons
253 decaying leptons. Leptons are clean signals to detect and trigger on.

254 Since the Higgs couples proportional to mass as already mentioned, decays to heavier
255 particles are favoured. The branching ratios of different Higgs decay modes are
256 shown as a function of m_H in Fig. 1.3. Approximately 60% of the time the Higgs
257 decays to a pair of b -quarks, the dominant decay mode since the Higgs is not massive
258 enough to decay to a pair of top quarks (recall from Table 1.1 $m_b = 4.18 \text{ GeV}$ and
259 $m_t = 173 \text{ GeV}$). The next heaviest fermions are the tau lepton and the c -quark,
260 decays to pairs of these particles happen approximately an order of magnitude less
261 often. Decays to pairs of vector bosons are via a virtual off shell Higgs only, since
262 the combined vector boson mass is greater than the Higgs mass. While the $H \rightarrow \gamma\gamma$
263 and $H \rightarrow Z\gamma$ branching ratios are very small compared with fermionic decay modes
264 (around 0.2% for $H \rightarrow \gamma\gamma$), these decay channels were instrumental in the initial
265 discovery of the Higgs due to the low level of background processes which mimic the
266 final state.

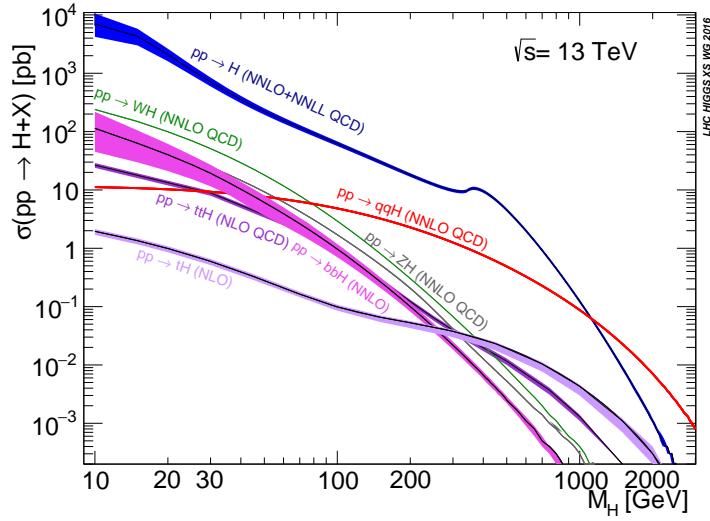


Figure 1.2: Higgs production cross sections as a function of Higgs mass m_H at $\sqrt{s} = 13$ TeV [24]. Uncertainties are shown in the shaded bands. At the Higgs mass $m_H = 125$ GeV, Higgs production is dominated by gluon-gluon fussion, vector boson fusion, and associated production with vector bosons.

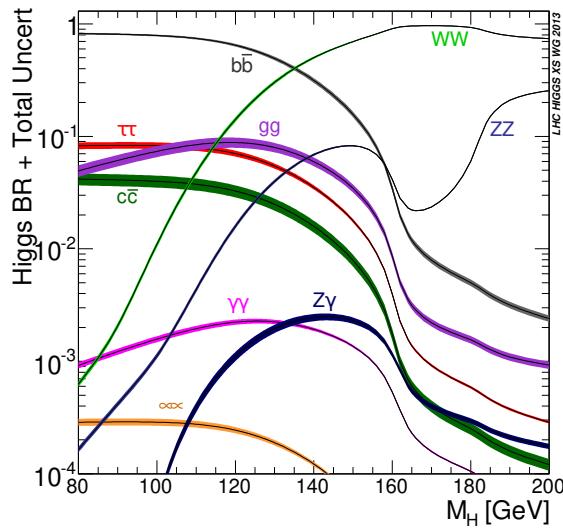


Figure 1.3: Higgs branching ratios as a function of Higgs mass m_H at $\sqrt{s} = 13$ TeV [24]. Uncertainties are shown in the shaded bands. At the Higgs mass $m_H = 125$ GeV, the Higgs predominantly decays to a pair of b -quarks, around 60% of the time. The subdominant decay mode is off shell to a pair of W bosons.

267 This thesis presents Higgs analyses using events with a Higgs produced in association
268 with vector boson, where the Higgs decays to a pair of b -quarks, i.e. $pp \rightarrow VH(bb)$.
269 The $H \rightarrow b\bar{b}$ decay mode directly probes the Higgs coupling to the second generation
270 fermions, and more specifically to bottom quark. This coupling was first observed in
271 2018 [25, 26]. Ongoing work measuring the coupling strengths, in particular in the
272 high energy regime, is the focus of the analysis presented in this thesis in Chapter 6.

²⁷³ **Chapter 2**

²⁷⁴ **The Large Hadron Collider and the
ATLAS Detector**

²⁷⁶ The Large Hadron Collider (LHC) at CERN has extended the frontiers of particle
²⁷⁷ physics through its unprecedeted energy and luminosity. The LHC accelerates
²⁷⁸ protons around a 27 km ring until they are travelling just 3 m s^{-1} slower than than
²⁷⁹ the speed of light, at which point they are made to collide. The protons travel round
²⁸⁰ the ring 11 thousand times per second in two concentric beams, which are guided by
²⁸¹ superconducting magnets cooled using liquid helium to -271.3°C (1.9 K). The beams
²⁸² travel in opposite directions and are crossed at four locations so that collisisons
²⁸³ between protons can take place. Around these collision points four specialised
²⁸⁴ detectors, ALICE [27], CMS [28], LHCb [29] and ATLAS [30], are located to capture
²⁸⁵ information about the products of the collisions.

²⁸⁶ The LHC is operated in *runs* during which beams of protons are actively being
²⁸⁷ circulated and collided. Between runs which there are periods of shutdown while the
²⁸⁸ accelerator and detector machinery is maintained and upgraded. In 2010, the LHC
²⁸⁹ collided proton bunches, each containing more than 10^{11} particles, 20 million times
²⁹⁰ per second, providing 7 TeV proton-proton collisions at instantaneous luminosities of
²⁹¹ up to $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. Run 2, which began in 2015, increased the the proton-
²⁹² proton collision energy to 13 TeV. The bunch spacing was also reduced, leading to
²⁹³ a collisison rate of 40 MHz. Over the course of Run 2 a total integrated luminosity
²⁹⁴ of 146.9 fb^{-1} was recorded. 2022 marked the beginning of Run 3 which, with a
²⁹⁵ higher center of mass energy and peak luminosity, is expected to culminate in the
²⁹⁶ approximate doubling of the dataset size.

include ac-
celerator
complex
info?

Period	Year	\sqrt{s} [TeV]	$\langle\mu\rangle$	Bunch spacing [ns]	Luminosity [$\text{cm}^{-2} \text{s}^{-1}$]
Run 1	2011–2012	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 2.1: Overview of the different LHC runs [31, 32]. The average number of interactions per bunch-crossing is denoted as $\langle\mu\rangle$ (see Section 2.1.3). Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.

297 2.1 Coordinate System & Collider Definitions

298 In Section 2.1.1, the coordinate system used at ATLAS is introduced. The parameter-
 299 isation used for the specifying the trajectory of charged particle tracks is described in
 300 Section 2.1.2, and definitions for some frequently occurring concepts and quantities
 301 is provided in Section 2.1.3.

302 **2.1.1 ATLAS Coordinate System**

303 The origin of the coordinate system used by ATLAS is at the nominal interaction
 304 point in the centre of the detector. As shown in Fig. 2.1, the z -axis points in the
 305 forward direction along the beam pipe, while the x -axis points from the interaction
 306 point to the centre of the LHC ring, and the y -axis points upwards. The transverse
 307 plane lies in x - y while the longitudinal plane lies along the z -axis. A cylindrical
 308 coordinate system with coordinates (r, ϕ) is used in the transverse plane, where r is
 309 the radius from the origin and ϕ is the azimuthal angle around the z -axis.

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

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]. \quad (2.1)$$

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

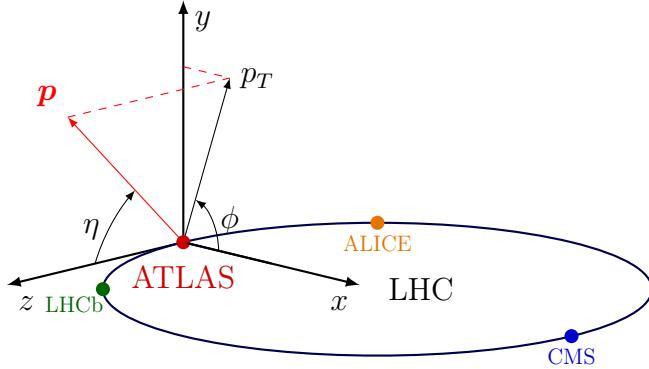


Figure 2.1: The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [33].

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

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (2.2)$$

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

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (2.3)$$

318 2.1.2 Track Parameterisation

319 The trajectories of charged particle tracks are specified using five parameters:
320 $(d_0, z_0, \phi, \theta, q/p)$. Transverse and longitudinal impact parameters (IP) d_0 and
321 z_0 specify the closest approach of the trajectory of a particle to the origin. The
322 transverse IP d_0 and longitudinal IP z_0 are measured with respect to the hard scatter
323 primary vertex (see Section 2.3.2). ϕ and θ are the azimuthal and polar angles
324 respectively, and q/p is the measured charge on the track¹ divided by the scalar
325 3-momentum. Fig. 2.2 shows each of these parameters diagrammatically.

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

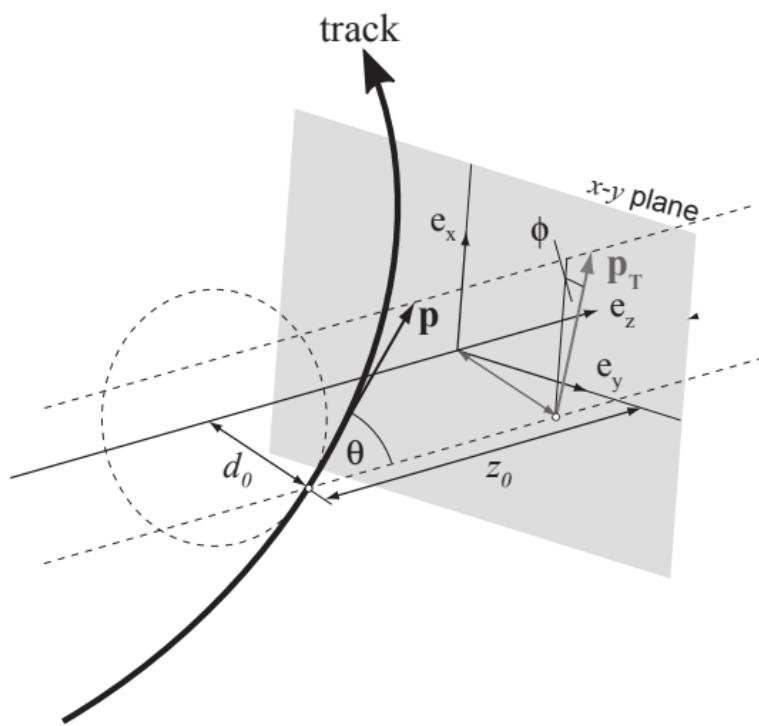


Figure 2.2: The track parameterisation used at the ATLAS detector. Five coordinates ($d_0, z_0, \phi, \theta, q/p$) are specified, defined at the point closest approach of the track to the nominal interaction point at the origin of the coordinate system.

326 Impact parameter significances are defined as the IP divided by its corresponding
327 uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. When used in flavour tagging
328 (see Chapter 3), track IP significances are lifetime signed according to the track's
329 direction with respect to the jet axis and the primary vertex [34].

330 2.1.3 Hadron Collider Definitions

331 Cross Section

332 Cross sections are of primary importance to experimental particle physics. The cross
333 section σ for a given process is closely related to the probability of an interaction
334 between two colliding particles, and is analogous to an effective cross-sectional area
335 of the particles. The cross section itself depends on the transition matrix element,
336 obtained using the Feynman rules of the theory which are derived using QFT, and a
337 phase space integral. At hadron colliders such as the LHC, the proton-proton cross
338 section can be factorised as

$$\sigma(pp \rightarrow X) \approx \text{PDFs} \cdot \text{partonic cross section.} \quad (2.4)$$

339 The partonic cross section can be calculated, while the parton distribution functions
340 (PDFs) have to be extracted from experimental results.

341 Luminosity

342 Total number of proton-proton collisions N is related to the total pp cross σ section
343 by the integrated luminosity L , as in

$$N = \sigma L = \sigma \int \mathcal{L} dt. \quad (2.5)$$

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

$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} = \frac{fn_1 n_2}{4\pi\sigma_x\sigma_y}, \quad (2.6)$$

346 where n_1 and n_2 are the number of protons in the colliding bunches, f is the bunch
 347 crossing frequency, and σ_x and σ_y are the rms width of the beam in the horizontal and
 348 vertical directions. At the LHC, the beams collide with a crossing angle $\phi \approx 285 \mu\text{rad}$
 349 to localise the interaction to a specific point in space. This reduces the effective
 350 luminosity.

351 The total luminosity recorded over the course of Run 2 is shown in Fig. 2.3. In total,
 139 fb^{-1} of usable physics data was collected over the three-year run.

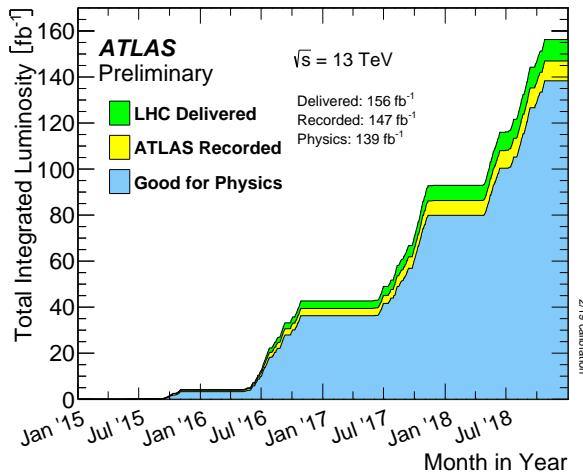


Figure 2.3: Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [32]. A total of 139 fb^{-1} of collision data is labelled as good for physics.

352

353 Pile-up

354 At the centre of the ATLAS detector, bunches of more than 10^{11} protons meet head
 355 on. Each bunch-crossing is called an *event*. There is generally at most one hard
 356 proton-proton scatter per event. Additional interactions are typically relatively soft
 357 and are known as *pile-up*. Pile-up complicates the reconstruction of the hard scatter
 358 event as results of the interactions of different proton-proton interactions have to
 359 be separated. Pile-up from interactions within the same bunch-crossing is known
 360 as *in-time* pile-up while residual signatures from previous bunch-crossings is known
 361 as *out-of-time* pile-up. The number of pile-up interactions is denoted μ , which is

³⁶² 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. 2.4.

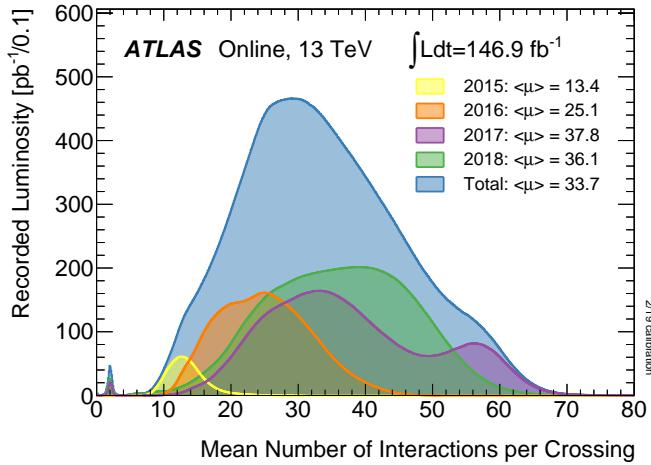


Figure 2.4: Average pile-up profiles measured by ATLAS during Run 2 [32]. During Run 3, even higher levels of pile-up are expected.

³⁶³

³⁶⁴ 2.2 The ATLAS Detector

³⁶⁵ The ATLAS² detector is a general-purpose particle detector based at the LHC.
³⁶⁶ The detector is made up of several specialised sub-detectors which are arranged
³⁶⁷ concentrically around the nominal interaction point at the centre of the detector,
³⁶⁸ covering nearly the entire solid angle around the collision point. In this section
³⁶⁹ a condensed overview of each sub-detector is given, in order of increasing radial
³⁷⁰ distance from the point of collision. The inner tracking detector is described in
³⁷¹ Section 2.2.1, the electromagnetic and hadronic calorimeters in Section 2.2.2, the
³⁷² muon spectrometer in Section 2.2.3, and the trigger in Section 2.2.4. More complete
³⁷³ information on the detector can be found in Ref. [30], or in the technical design
³⁷⁴ reports (TDRs) of the individual sub-detectors.

²A Toroidal Lhc ApparatuS.

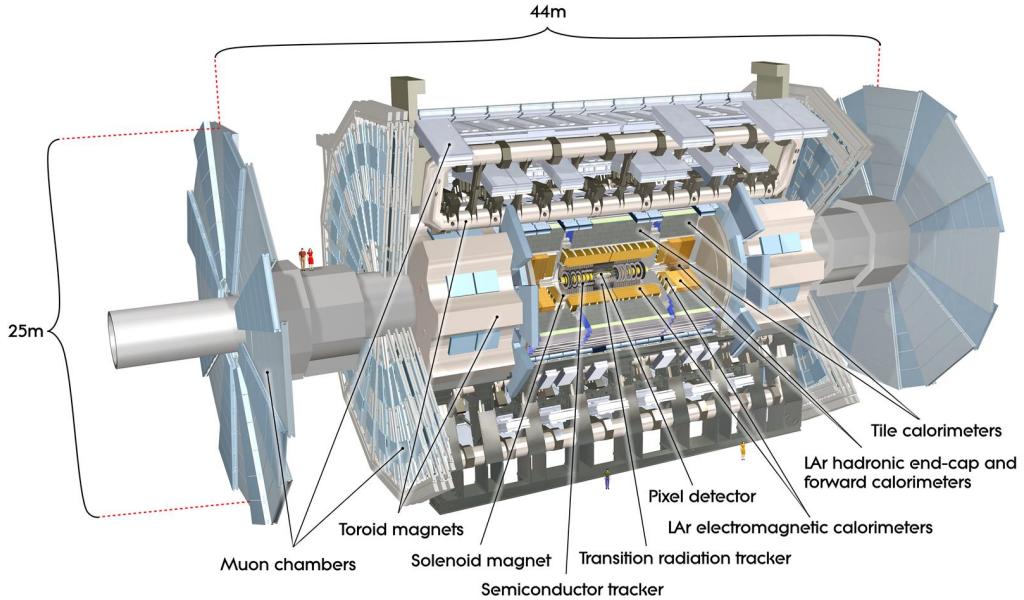


Figure 2.5: A 3D model of the entire ATLAS detector. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in layers around the nominal interaction point.

375 2.2.1 Inner Detector

376 The inner-detector system (ID) provides high-resolution charged particle trajectory
 377 tracking in the range $|\eta| < 2.5$. The ID is immersed in a 2 T axial magnetic field,
 378 produced by a superconducting solenoidal magnet, which enables the measurement
 379 of particle momentum and charge. After Run 3, the ID will be replaced by the
 380 ITk [35, 36].

381 The inner detector is made up of several sub-systems, shown in Figs. 2.6 and 2.7. The
 382 high-granularity silicon pixel detector covers the vertex region and typically provides
 383 four spacepoint measurements per track. It is followed by the silicon microstrip
 384 tracker (SCT), which usually provides a further four spacepoint measurements per
 385 track. These silicon detectors are complemented by the Transition Radiation Tracker
 386 (TRT), which enables radially extended track reconstruction up to $|\eta| = 2.0$.

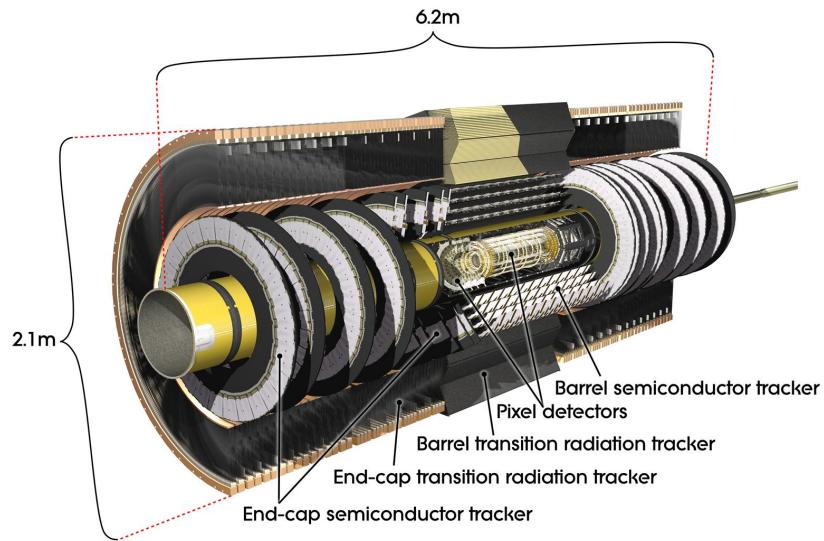


Figure 2.6: A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks.

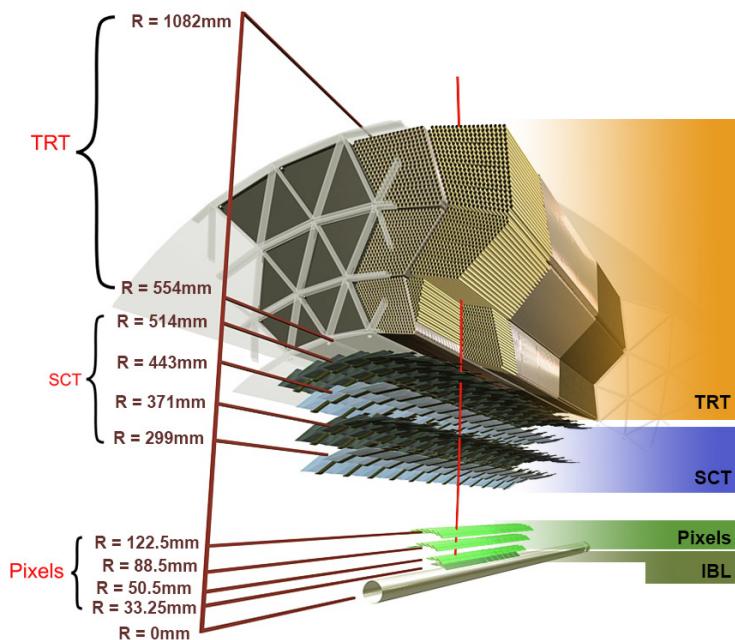


Figure 2.7: A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown.

387 Pixel Detector

388 The silicon pixel detector is comprised of four cylindrical barrels at increasing radii
389 from the beamline, and four disks on each side. The innermost barrel layer is the
390 insertable B-layer (IBL), which was installed before Run 2 [37,38] and lies just 33 mm
391 from the beam axis. The second-to-innermost layer is often referred to as the B-layer.
392 The specification of the pixel detector determines the impact parameter resolution
393 and the ability to reconstruct primary and secondary vertices. The detector is
394 required to have a high granularity (i.e. resolution) to maintain the low occupancy
395 required to resolve nearby particles. Individual pixels are 50 μm in the transverse
396 direction $R\phi$ and 400 μm in the longitudinal z direction (250 μm for the IBL). Cluster
397 positions have a resolution of approximately 10 μm in $R\phi$ and 100 μm in z .

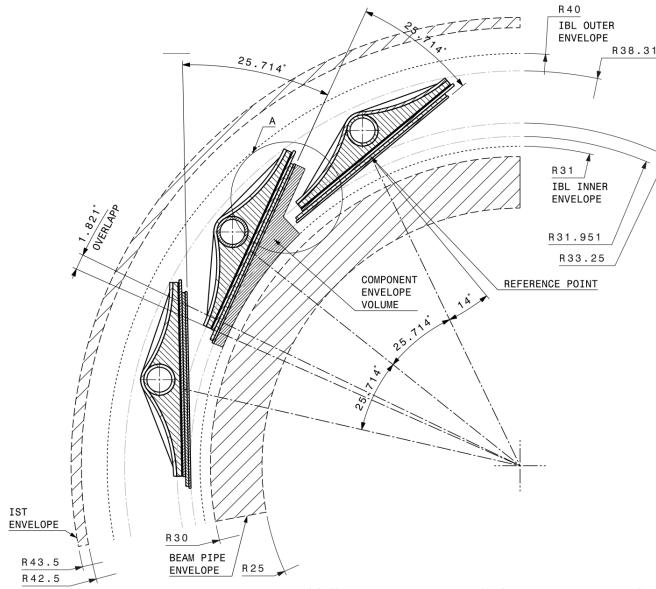


Figure 2.8: A schematic cross-sectional view of the ATLAS IBL.

398 Semi-Conductor Tracker (SCT)

399 The SCT is made up of four concentric barrel layers in the central region, and nine
400 disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,
401 with a small stereo angle (20 mrad) between the two layers enabling the z -coordinate
402 measurement from a pair of strip measurements. The SCT typically provides four

403 precision spacepoint measurements (eight strip measurements) per track in the barrel
404 region. These have intrinsic uncertainties of $17\text{ }\mu\text{m}$ in the transverse direction $R\phi$, and
405 $580\text{ }\mu\text{m}$ in the longitudinal direction z . The measurements provide a key contribution
406 to the measurement of charged particle momentum and impact parameter, along
407 with vertex position. Charge-particle tracks can be distinguished if separated by
408 more than $\sim 200\text{ }\mu\text{m}$. Hits are registered as binary signals if the pulse height in a
409 channel exceeds a certain threshold.

410 **Transition Radiation Tracker (TRT)**

411 The TRT is a straw-tube tracker which complements the higher-resolution silicon-
412 based tracks by offering a larger number of hits per track (typically around 30) and
413 a long lever arm, which aids the accurate measurement of particle momentum. It is
414 made up of approximately 300 000 drift tubes with a diameter of 4 mm which are
415 filled with xenon gas. The walls of each tube are electrically charged, and a thin
416 conducting wire runs along the center. When a charged particle traverses a tube, it
417 ionises the xenon and the resulting liberated electrons drift along the electric field
418 to the wire, where an associated charge is registered. In the barrel the straws run
419 parallel to the z -axis and therefore the TRT only provides tracking information in
420 $R\phi$. Straws are arranged radially in the end-caps. The resulting two-dimensional
421 spacepoints have a resolution of approximately $120\text{ }\mu\text{m}$. The spaces between the
422 straws are filled with a polymer which encourages the emission of transition radiation,
423 aiding electron identification.

424 **2.2.2 Calorimeters**

425 The calorimeter system measures the energy of incident particles over the range
426 $|\eta| < 4.9$. There are two main sub-systems: the electromagnetic calorimeter (ECal),
427 which focuses on the measurement of electrons and photons, and the hadronic
428 calorimeter (HCal), which measures the energy of hadrons. Upon entering the
429 calorimeter, incident particles will interact with the detector material to produce a
430 shower of secondary particles with reduced energies. The charge deposited in this
431 process is measured to reconstruct the energy of the initial incident particle. The
432 two calorimeter sub-systems must provide strong containment of showering

⁴³³ particles to prevent punch-through of EM and non-muon particles to the HCal and
⁴³⁴ muon system respectively.

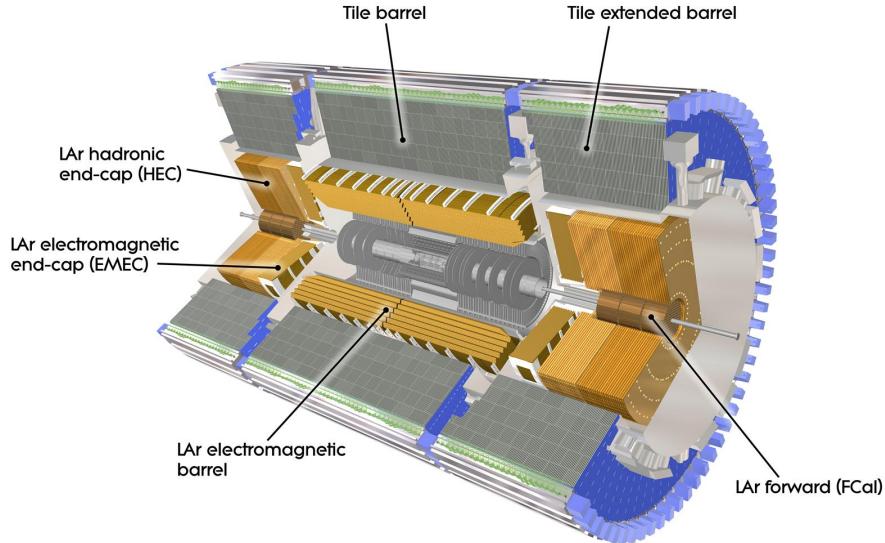


Figure 2.9: The ATLAS calorimeters. The ECal (orange) and HCal (grey, brown).

⁴³⁵ Liquid Argon (LAr) Electromagnetic Calorimeter

⁴³⁶ The more granular lead/liquid-argon ECal covers the region $|\eta| < 3.2$ and is split
⁴³⁷ into barrel (covering $|\eta| < 1.475$) and end-cap (covering $1.375 < |\eta| < 3.2$) regions.
⁴³⁸ EM calorimetry works by encouraging electrons and photons to interact with electric-
⁴³⁹ ally charged particles in detector material via bremsstrahlung ($e \rightarrow e\gamma$) and pair
⁴⁴⁰ production ($\gamma \rightarrow e^+e^-$). The EM calorimeter uses lead absorber plates to initiate
⁴⁴¹ EM showers, resulting in secondary particles which ionise the surrounding liquid
⁴⁴² argon. The charge is collected on copper electrodes and read out. The accordion
⁴⁴³ geometry of the ECal allows for a full coverage in ϕ without any azimuthal cracks.

⁴⁴⁴ Hadronic Tile Calorimeter

⁴⁴⁵ In the central barrel region with $|\eta| < 1.7$, the HCal uses a tile calorimeter with
⁴⁴⁶ steel as an absorbing material, and scintillating tiles as the active material. Two
⁴⁴⁷ copper/liquid-argon calorimeter end-caps are also used. Incident hadrons interact

⁴⁴⁸ via the strong and electromagnetic forces with the absorber material, mainly loosing
⁴⁴⁹ energy due to multiple inelastic nuclear collisions. The active material captures the
⁴⁵⁰ resulting electrons and photons to measure the energy of the incident hadron.

⁴⁵¹ 2.2.3 Muon Spectrometer

⁴⁵² Due to their higher mass, muons easily pass unimpeded through the ID and calorimeters
⁴⁵³ and therefore require specialised detectors for their measurement. The Muon
⁴⁵⁴ Spectrometer (MS) is made up of dedicated tracking and triggering hardware. The
⁴⁵⁵ precision tracking system uses three layers of monitored drift tubes with a barrel
⁴⁵⁶ region covering $|\eta| < 1.2$ and end-caps covering $1 < |\eta| < 2.7$. The inner layers of
⁴⁵⁷ the end-caps use cathode strip chambers to better cope with the high occupancy
⁴⁵⁸ in the forward region. Precision tracking resolution is approximately $50 \mu\text{m}$. The
⁴⁵⁹ trigger system is comprised of resistive plate chambers in the barrel region covering
⁴⁶⁰ $|\eta| < 1.0$ and thin gap chambers in the end-cap regions covering $1 < |\eta| < 2.4$. A set
⁴⁶¹ of three superconducting air-core toroidal magnets, each made up of eight coils, is
⁴⁶² used in each of the barrel and end-caps to deflect the muons as they pass through
⁴⁶³ the MS, allowing their momentum and charge to be measured from the direction
⁴⁶⁴ and magnitude of curvature. The toroidal magnets generate a field which is largely
⁴⁶⁵ orthogonal to the muon trajectories which allows for maximum deflection.

⁴⁶⁶ 2.2.4 The Trigger

⁴⁶⁷ The 2.5 ns bunch spacing used over the course of Run 2 corresponds to a bunch-
⁴⁶⁸ crossing or event rate of 40 MHz (see Table 2.1). If the full information for the
⁴⁶⁹ detector was written out for each event, this would correspond to the generation
⁴⁷⁰ of 60 TB of data each second. This is more than can be feasibly processed and
⁴⁷¹ stored, requiring the use of a trigger system which quickly makes a decision about
⁴⁷² whether or not an event is potentially interesting and should be kept for further
⁴⁷³ analysis. The trigger system is comprised of two levels which search for signs of
⁴⁷⁴ electrons, muons, photons, taus and jets, as well as events with high total or missing
⁴⁷⁵ transverse energy. The hardware-based Level-1 (L1) trigger uses coarse information
⁴⁷⁶ from the calorimeters and MS to accept events at a rate of 100 kHz within $2.5 \mu\text{s}$
⁴⁷⁷ of the event. After the L1 trigger, the software-based High Level Trigger (HLT)

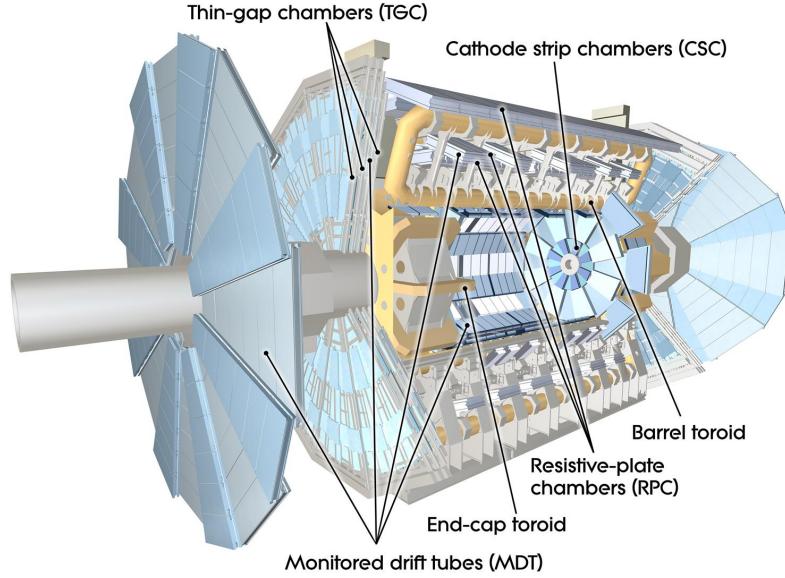


Figure 2.10: The ATLAS muon spectrometer.

478 makes use of 40 000 CPU cores to make a final selection on surviving events within a
 479 few hundred milliseconds. The final event read-out rate is approximately 1.2 kHz,
 480 corresponding 1.2 GB s^{-1} of permanent data storage. More information is provided
 481 in [39].

482 2.3 Reconstructed Physics Objects

483 Event reconstruction is the process of analysing the raw signals from the detector to
 484 determine the type and properties of particles present in an event. The reconstructed
 485 event provides information about the underlying physics process that led to the
 486 observable final state. Events passing the trigger selection (described in Section 2.2.4)
 487 undergo full offline reconstruction, which makes use of the full information from
 488 the detector. Reconstruction and analysis of events relies on the extensive ATLAS
 489 software stack, see Ref. [40] for more information.

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

492 2.3.1 Tracks

493 The trajectories of charged particles are reconstructed as tracks from the energy
494 depositions (called *hits*) left by the particles as they traverse the sensitive elements
495 of the inner detector. Tracks are used for a variety of downstream applications,
496 including vertexing and jet tagging. A comprehensive introduction to ATLAS
497 tracking is available in [41], while specific optimisations for dense environments are
498 detailed in [42, 43]. An overview of track reconstruction is given below.

499 Space-point Formation (Clustering)

500 When a charged particle traverses a pixel layer, charge is typically collected in more
501 than one pixel. This is due to the incident angle of the particles with respect to the
502 sensor, and also the drift of electrons between sensors caused by the magnetic field.
503 Clusters (also called *hits* or space-points) are formed by clustering neighbouring pixel
504 cells and estimating locations of space-points using the shape and energy distribution
505 of the clusters.

506 Track Finding

507 Space-points are used to build track seeds. These are groups of three hits which
508 are geometrically compatible with being part of a track segment. A combinatorial
509 Kalman filter (KF) is used to build track candidates by extending track seeds. The
510 filter can create multiple track candidates per seed, with bifurcations along the
511 track occurring when more than one compatible space-point exists on a given layer.
512 In this way, the KF creates an excessive number of *track candidates*, which are
513 only required to satisfy basic quality requirements. Track candidates are allowed to
514 reuse or *share* hits freely (a single hit may be used by multiple track candidates).
515 Typically, the presence of shared hits is a predictor of a bad track due to the high
516 granularity of the ATLAS tracking detectors. At this stage, there is a large number
517 of incorrect hits assigned to otherwise good tracks, and additionally large number of
518 *fake* tracks, which are comprised of a majority of wrong hits and do not correspond
519 to the trajectory of any one physical particle (fake tracks are defined as those where
520 the majority of associated hits do not originate from one single truth particle, see

521 Eq. (4.5)). The low quality of tracks at this stage necessitates an ambiguity solving
522 step, in which candidates are cleaned, and the highest quality track are selected.

523 Ambiguity Solving

524 Ambiguity solving was introduced as part of the ATLAS New Tracking effort (NEWT)
525 [41], and is intended to improve track reconstruction performance in challenging dense
526 environments. In the ambiguity solver, track candidates are processed individually
527 in descending order of a track score. The track score quantifies the likelihood of the
528 track corresponding to the trajectory of a real particle. Scoring uses a number of
529 signals, including the number and positions of hits (preferring hits in more precise
530 regions of the detector), the transverse momentum of the track and the track fit
531 quality. The track fit quality describes the quality of the track as the χ^2 divided
532 by the number degrees of freedom on the track. A preference for high transverse
533 momentum tracks promotes the successful reconstruction of the more physically
534 interesting energetic particles, and suppresses the large number of wrong hits assigned
535 to low momentum tracks.

536 During the processing of a given highest-scoring track candidate, the track is cleaned
537 (whereby problematic hits are removed), and, if the cleaning is successful, a full
538 resolution fit is performed. If the track has reached this stage without rejection by
539 passing various quality regiments, it is re-scored and returned to the list of track
540 candidates. If the same track is then processed again without requiring modification,
541 it is added to the final track collection. Track candidates that fall beyond a certain
542 quality cut are rejected. This cut does allow the possibility of a track passing through
543 the ambiguity solver with a small number of shared hits.

list shared
hit cut

544 Neural Network Cluster Splitting

545 As part of track cleaning, shared hits are classified by a Neural Network (NN)
546 to determine if they are compatible with the characteristic features of a merged
547 cluster [42, 44]. A merged cluster is one which has originated from more than
548 one incident particle. The corresponding reconstructed cluster is made up of a
549 combination of energy deposits from more than one particle, which have become
550 merged due to the closeness of the associated particles and the limited resolution of

551 the detector. While in general this event is rare, it is common for clusters to become
 552 merged in dense environments, as discussed in Section 3.1. If the cluster is predicted
 553 to be merged it is labelled as being freely shareable, or *split*. Hits classified as split
 554 by the cluster splitting NN are allowed to be shared freely. Hits not compatible with
 555 the merged hypothesis can still be shared by a limited number of tracks, but come
 556 with a penalty for the track which may hinder its acceptance into the final track
 collection.

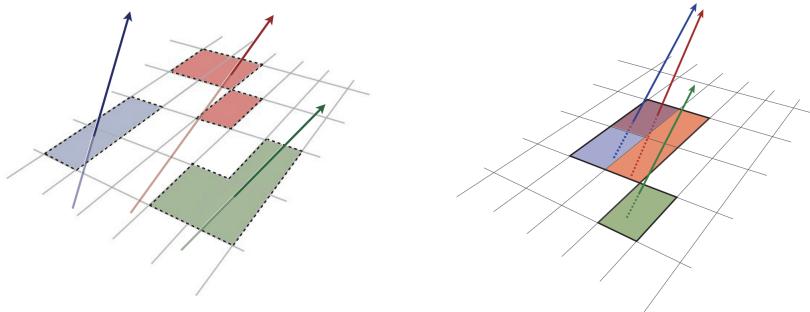


Figure 2.11: Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster.

557

558 2.3.2 Vertices

559 Groups of reconstructed tracks can be examined to determine whether the particles
 560 originated from a common spatial point of origin. This occurs when a particle
 561 decays or radiates. Vertex reconstruction is made up of two stages. First, vertex
 562 finding takes place, which is the process of grouping tracks into compatible vertices.
 563 Second, vertex fitting combines information from compatible tracks to reconstruct
 564 the physical properties of the vertex, such as mass and position.

565 Primary Vertices

566 Each proton-proton interaction happens at a *primary vertex* (PV). Primary vertices
 567 are iteratively reconstructed with tracks using the iterative vertex finder (IVF) [45].
 568 In Run 3, the IVF will be replaced with an adaptive multi-vertex finder (AMVF) [46].

569 The *hard scatter vertex* of an event is chosen as the primary vertex whose associated
570 tracks have the largest sum of transverse momentum squared, Σp_T^2 .

571 **Secondary Vertices**

572 Secondary vertices (SV) occur when a particle radiates or decays at a sufficient
573 distance from the primary vertex to be resolved from the primary vertex (see
574 Section 3.1.1). Two widely used secondary vertexing tools are used within ATLAS:
575 SV1 and JetFitter [47]. Each attempts to reconstruct secondary vertices inside a jet
576 given the tracks associated to that jet (see Section 2.3.3 for more information about
577 track association). SV1 by design attempts to reconstruct only a single inclusive
578 vertex per jet. This inclusive vertex groups all b -hadron decay products, including
579 tracks from the b -hadron decay itself and tracks from $b \rightarrow c$ decays. The second tool,
580 JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary
581 vertices from b -hadron decays are reconstructed separately to tertiary vertices from
582 $b \rightarrow c$ decay chains.

583 **2.3.3 Jets**

584 Jets are an aggregate reconstructed object corresponding to a spray of collimated
585 stable particles which results from a decay chain of an unstable progenitor particle.
586 Jets are built by clustering constituent objects (e.g. tracks or calorimeter clusters)
587 using a jet finding algorithm, for example the anti- k_t algorithm [48].

588 **Particle Flow Jets**

589 Particle-flow (PFlow) jets are reconstructed from particle-flow objects [49] using
590 the anti- k_t algorithm with a radius parameter of 0.4. Particle-flow objects integrate
591 information from both the ID and the calorimeters, improving the energy resolution
592 at high transverse momenta and reducing pileup contamination. The PFlow jet
593 energy scale is calibrated according to Ref. [50].

594 Tracks are associated to jets using a ΔR association cone, the width of which
595 decreases as a function of jet p_T , with a maximum cone size of $\Delta R \approx 0.45$ for jets

596 with $p_T = 20 \text{ GeV}$ and minimum cone size of $\Delta R \approx 0.25$ for jets with $p_T > 200 \text{ GeV}$.
597 If a track is within the association cones of more than one jet, it is assigned to the
598 jet which has a smaller $\Delta R(\text{track}, \text{jet})$.

599 Jet flavour labels are assigned according to the presence of a truth hadron within
600 $\Delta R(\text{hadron}, \text{jet}) < 0.3$ of the jet axis. If a b -hadron is found the jet is labelled a b -jet.
601 In the absence of a b -hadron, if a c -hadron is found the jet is called a c -jet. If no b -
602 or c -hadrons are found, but a τ is found in the jet, it is labelled as a τ -jet, else it is
603 labelled as a light-jet.

604 PFlow jets are used to train the algorithms discussed in Chapter 4 and Chapter 5.

605 Track-jets

606 Track-jets are built by clustering tracks using the anti- k_t clustering algorithm and
607 are used in the analysis described in Chapter 6. The radius parameter is allowed to
608 vary with transverse momentum such that a broader cone (up to $R = 0.4$) is used for
609 high- p_T track-jets and a narrower cone (down to $R = 0.02$) for high- p_T track-jets.
610 Truth flavour labels for track-jets are derived using the same $\Delta R(\text{hadron}, \text{jet}) < 0.3$
611 matching scheme as for PFlow jets.

612 Large- R Jets

613 Large- R jets have a radius parameter $R = 1.0$ and are built by clustering topological
614 calorimeter clusters using the anti- k_t algorithm [51]. The large radius parameter
615 is especially useful for containing the decay products of a boosted Higgs boson, as
616 discussed in Chapter 6.

617 2.3.4 Leptons

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

621 **Electrons**

622 Electrons candidates are reconstructed by matching PV-compatible inner detector
 623 tracks to calorimeter clusters. The track-cluster matching criteria takes into account
 624 the significant energy loss of the electron due to bremsstrahlung. If a match is
 625 found, a refit of the track is performed using the Gaussian Sum Filter (GSF) [52],
 626 which better handles trajectory reconstruction in the presence of bremsstrahlung.
 627 Various identification criteria are then applied using a likelihood-based (LH) method
 628 to the candidates to improve purity. These include requirements on the track quality
 629 and cluster matching, the shape of shower in the ECal, leakage into the HCal, and
 630 the amount of transition radiation detected in the TRT. A full description can be
 obtained in Ref. [53].

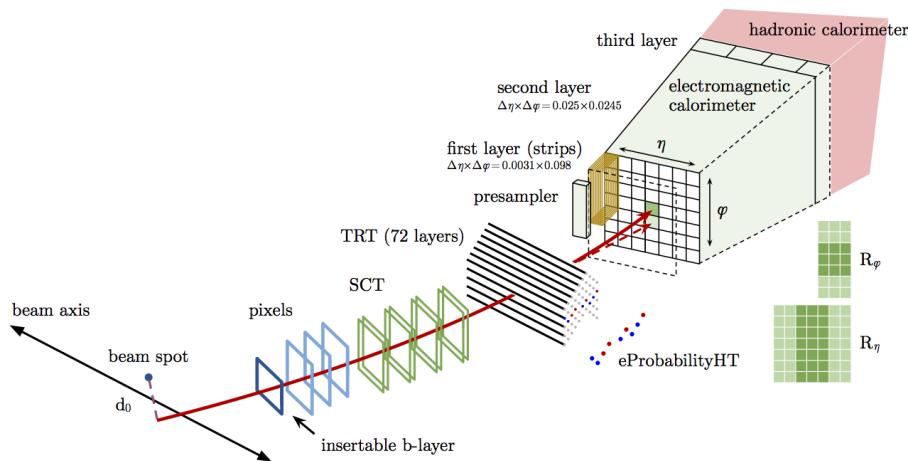


Figure 2.12: A sketch of electron reconstruction using the ATLAS detector [53]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.

631

632 **Muons**

633 Muon reconstruction primarily makes use of the dedicated MS (see Section 2.2.3), but
 634 also relies on tracks from the ID and the presence of characteristic signatures in the
 635 calorimeters. Muon tracks are reconstructed by connect straight-line track segments,
 636 which are identified via a Hough transform, and combined into a approximately
 637 parabolic trajectory. Finally, a global χ^2 fit is performed, taking into account possible

638 interactions between the muon and the detector material. A reconstructed muon is
639 called *combined* if it completes successful matching to an ID track. Combined muons
640 undergo a further fit with the combined ID and MS hits, with the energy loss due to
641 the traversal of the calorimeters being taking into account.

642 **2.3.5 Missing Transverse Momentum**

643 worth adding?

644

Chapter 3

645

Tracking and b -tagging

646 Many ATLAS analyses rely on b -tagging, which is a method of selecting jets ini-
647 stantiated by b -quarks (b -jets) and rejecting jets created from other quarks (c and
648 light flavours u , d , s). These b -tagging algorithms work by discriminating for the
649 unique signatures of b -jets discussed in Section 3.1. The various tagging algorithms
650 ultimately take as their input information about the reconstructed jet and any
651 associated tracks. Successful b -tagging relies therefore on the efficient and accurate
652 reconstruction of tracks, and especially those tracks corresponding to the products
653 of b -hadron decays.

654 Historically a two tiered approach has been taken, in which so called *low-level* taggers
655 take as inputs information about the jet and associated tracks, and then outputs of
656 several low-level taggers are fed into a *high-level* tagger which uses a multivariate
657 approach to discriminate between jet flavours. , as these tracks are used as the
658 primary inputs to vertex reconstruction algorithms and jet tagging algorithms (for
659 more information see Chapter 5).

660 This chapter summarises the challenges facing tracking and b -tagging at high trans-
661 verse momentum with an investigation into track reconstruction performance in
662 Section 3.1. Some preliminary investigations into improving tracking in this regime
663 are investigated in Section 3.2.

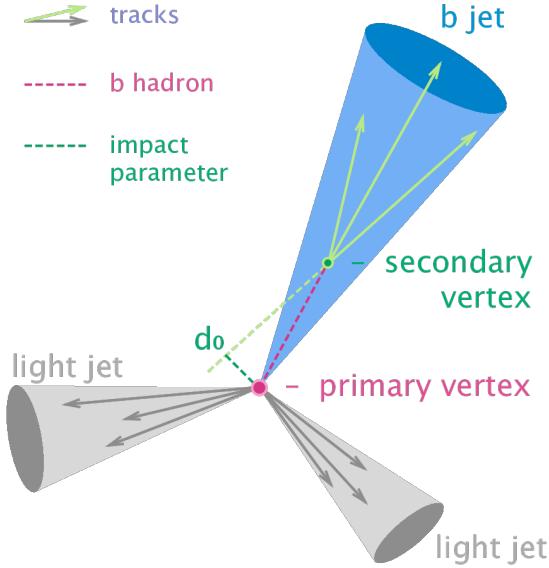


Figure 3.1: Diagram of a typical b -jet (blue) which has been produced in an event alongside two light jets (grey). The b -hadron has travelled a significant distance (pink dashed line) from the primary interaction point (pink dot) before its decay. The large transverse impact parameter d_0 is a characteristic property of the trajectories of b -hadron decay products.

664 3.1 b -hadron Reconstruction

665 This section outlines the typical detector signature of a b -hadron in Section 3.1.1
 666 and discusses some associated reconstruction difficulties in Section 3.1.2.

667 3.1.1 Decay Topology

668 b -hadrons are quasi-stable bound states of a bottom quark and one or more lighter
 669 quarks. Collectively, these are the b -mesons (e.g. $B^+ = u\bar{b}$, $B^0 = d\bar{b}$) and baryons
 670 with (e.g. $\Lambda_b^0 = udb$). After a b -quark is produced as the result of some proton-proton
 671 collision, they quickly hadronise. The hadronisation process is hard – around 70-80%
 672 of the b -quark’s momentum is passed to the b -hadron, with the rest being radiated
 673 as prompt hadronisation or fragmentation particles. See Ref. [54] for a more in
 674 depth discussion on hadronisation and the closely related process of fragmentation.
 675 Henceforth the combined hadronisation and fragmentation products will be referred
 676 to collectively as fragmentation.

677 *b*-hadrons are interesting objects of study due to their relatively long proper life-
678 times $\tau \approx 1.5$ ps. Early studies showed that *b*-hadrons did not couple strongly to
679 light-flavour quarks [55]. The lifetime of *b*-hadrons is therefore approximately deter-
680 mined only by a single CKM matrix element V_{cb} (see Section 1.1.3). This lifetime
681 corresponds to a proper decay length $c\tau \approx 450$ μm . In the rest frame of the detector,
682 the typical *b*-hadron travels a distance

$$d = \gamma\beta c\tau \approx \gamma c\tau \quad (3.1)$$

683 before decaying, where in the high energy limit $\gamma = E_b/m_b$ and $\beta = v/c = 1$. For
684 a 1 TeV *b*-hadron, this gives $d \approx 90$ mm – well beyond the radius of the first pixel
685 layer (the IBL) at a radius of 33 mm (see Fig. 3.2). This significant displacement is
686 characteristic of *b*-jets and makes it possible to reconstruct secondary vertices at the
687 *b*-hadron decay point.

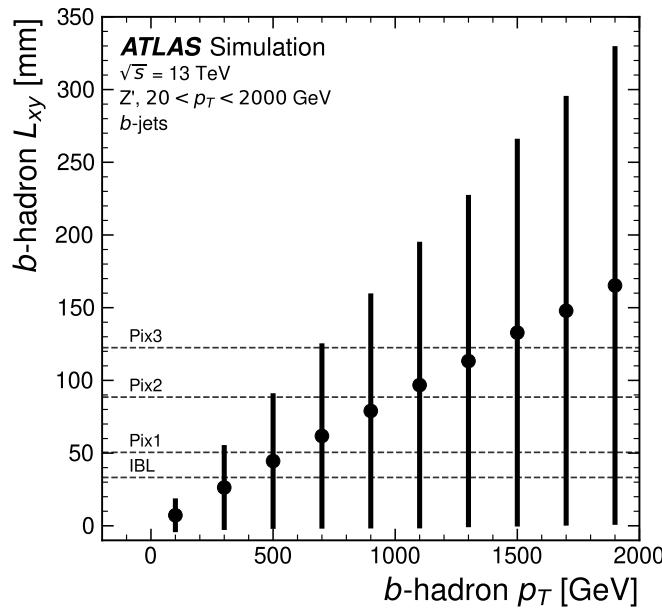


Figure 3.2: The truth *b*-hadron decay radius L_{xy} as a function of truth transverse momentum p_T for reconstructed *b*-jets in Z' events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.

688 *b*-hadrons decay weakly to on average four or five collimated stable particles. These
689 particles, along with any fragmentation particles, are reconstructed in the detector as
690 a jet. A *b*-jet has several characteristic features which differentiate it from light-jets.

These features stem from the significant displacement of the *b*-hadron that can occur in high transverse momentum *b*-jets. Associated tracks and SVs can have a large transverse impact parameter d_0 as a result of the *b*-hadron displacement (as shown in Fig. 3.1). Additionally, 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. While the multiplicity of the fragmentation products increases with the *b*-hadron p_T , the multiplicity of the products of the weak decay is unaffected.

3.1.2 Challenges Facing *b*-hadron Reconstruction

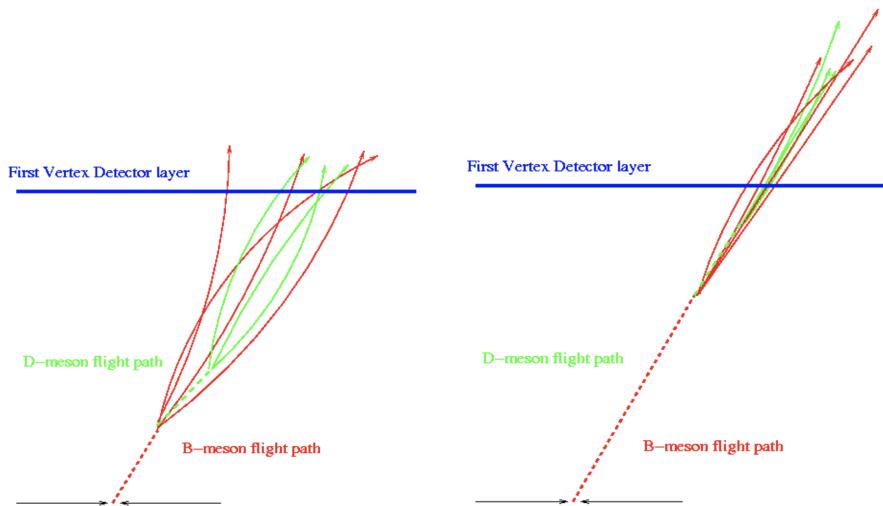


Figure 3.3: At lower p_T (left) the decay length of the *b*-hadron is reduced, and the resulting decay tracks are less collimated. At higher p_T (right) the *b*-hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge before reaching detector elements. As a result, the ID may be unable to resolve charged depositions from different particles, instead reconstructing merged clusters.

As discussed, a necessary requirement for successful jet *b*-tagging is the efficient and accurate reconstruction of the charged particle trajectories in the jet. For high p_T jets ($p_T > 200$ GeV) this task becomes difficult due to a combination of effects. As the jet energy increases, the multiplicity of tracks in the jet increases due to the presence of additional fragmentation particles. Fragmentation and weak decay products also become increasingly collimated as their inherited transverse momentum increases. Together, these two effects lead to a high density of charged particles in the jet core, which, given the finite resolution of the detector, makes reconstruction

707 difficult. At high energies, the increased decay length of b -hadrons (and c -hadrons)
 708 means that decay products have less of an opportunity to diverge before reaching
 709 the first tracking layers of the detector (shown in Fig. 3.3). If the weak decay takes
 710 place close enough to a detector layer, or if the particles are otherwise sufficiently
 711 collimated, charge deposits left by nearby particles may not be resolved individually,
 712 instead being reconstructed as merged clusters. As discussed in Section 2.3.1, merged
 713 clusters are generally rare, and so shared hits generally predict bad tracks and
 714 are correspondingly penalised during track reconstruction. However, in the core of
 715 high p_T b -jets the density of particles is high enough that the probability of cluster
 716 merging increases dramatically. Successful reconstruction of such tracks requires
 717 the presence of shared hits, but the presence of these can paradoxically end up
 718 impairing the successfully reconstruction of the track. Furthermore, decays may also
 719 take place inside the tracking detectors themselves, which at best leads to missing
 720 measurements on the most sensitive detector layers, and at worst can lead to wrong
 721 inner layer hits being added to displaced tracks, since the reconstruction process
 722 penalises tracks without inner layer hits. The combination of effects described above
 723 makes reconstructing tracks in the core of high p_T b -jets particularly challenging.

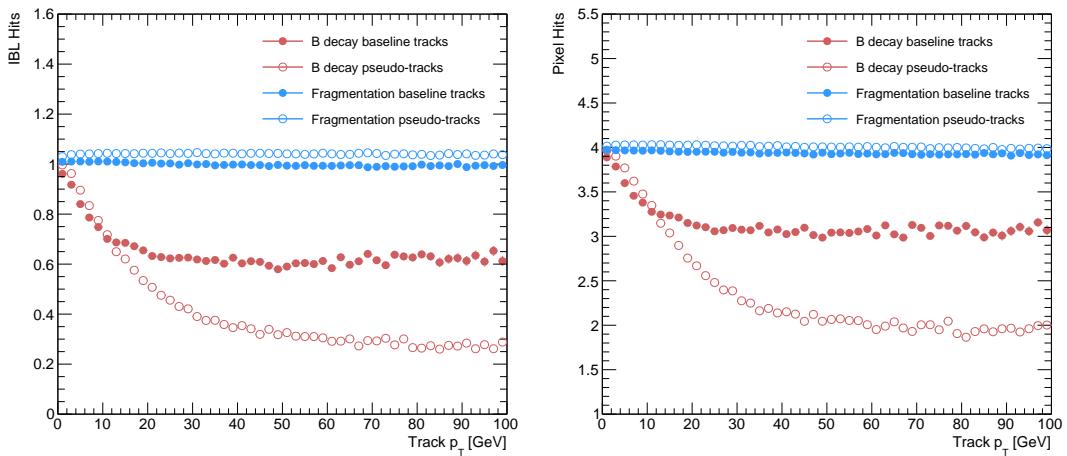


Figure 3.4: Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the p_T of the reconstructed track. Tracks from the weak decay of the b -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. For each of these, standard tracks and pseudo-tracks are plotted. Hit multiplicities on the pseudo-tracks at high p_T due to the increased flight of the b -hadron. The baseline tracks have more hits than the pseudo-tracks, indicating that they are being incorrectly assigned additional hits.

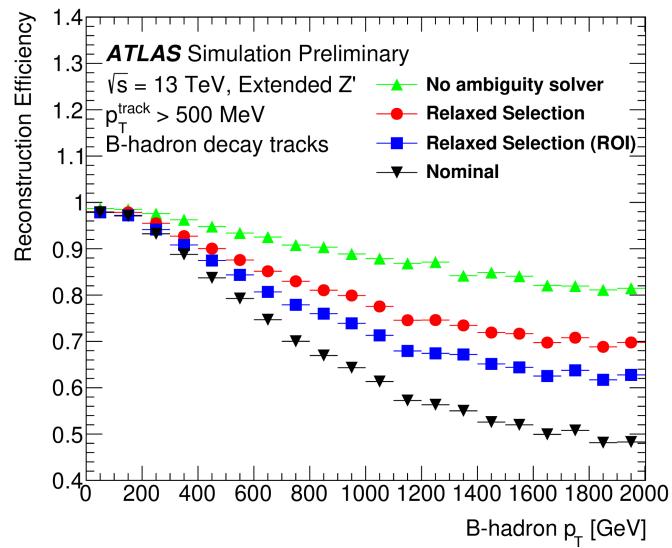


Figure 3.5: *b*-hadron decay track reconstruction efficiency as a function of truth *b*-hadron p_T [56]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- p_T *b*-hadrons, 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.

724 The above effects create two related, but distinct problems for b -tagging. The first
725 part is a drop in track reconstruction efficiency. As mentioned, tracks originating
726 from high energy b -hadron decay products can have a high rate of shared hits due
727 to the number of particles present in a high p_T b -jet and their relative collimation.
728 Additionally, tracks may be missing hits on the inner layers of the detector in the
729 case of displaced decays. The presence of shared and missing hits reduces a track's
730 score in the ambiguity solver meaning that higher ranking, but potentially worse,
731 track candidates are processed first and take ownership of the hits. This can make
732 it difficult for otherwise reasonable b -hadron decay tracks to meet the ambiguity
733 solver's stringent track quality requirements, leading to their rejection at this stage.
734 this is shown in Fig. 3.5.

735 The second part of the problem is that, due to the high multiplicity of clusters
736 available for assignment in the vicinity of the typical high energy b -hadron decay
737 track, and also given the strong positive bias of the ambiguity solver towards those
738 tracks with precise pixel measurements (especially the innermost IBL measurement),
739 many b -hadron decay tracks are assigned incorrect inner layer hits. This is only a
740 problem for those decay products which were produced inside the pixel detector
741 as a result of a long-flying b -hadron, and so do not have a correct hit available
742 for assignment. The incorrect hits may skew the parameters of the track, which
743 can in turn mislead the downstream b -tagging algorithms. In particular, b -tagging
744 algorithms rely heavily on the transverse impact parameter significance $s(d_0)$ of the
745 track. The quality of this measurement is expected to be adversely affected by wrong
746 inner-layer hits on the track.

747 The combination of reduced reconstruction efficiency and incorrectly assigned hits is
748 thought to be the cause of the observed drop in b -tagging efficiency at high energies,
749 however further study is required to determine which effect may dominate.

cite sebs
study show-
ing they are
approx simi-
lar impacts?

750 3.2 Investigations into High p_T b -hadron Tracking

751 In Section 3.2.1 pseudotracks, a key tool for studying the ideal tracking performance
752 of the ATLAS detector, are used to study the baseline hit-to-track association in
753 the dense cores of high- p_T b -jets. Section 3.2.2 details a study which investigated
754 modifying the global track fitter to improve reconstruction performance in this

regime. Not detailed here is an investigation into the bcut + refit method. Shown not to be promising, as, even though an improvement in the *b*-hadron decay track efficiency was observed, the corresponding increase in the fake track was shown to be unacceptable.

3.2.1 Pseudotracking

Pseudo-tracking uses Monte Carlo truth information to group together all the hits left by each truth particle. Each group of hits which passes basic quality requirements is directly used in a full resolution track fit. If the track fit is successful, a “pseudo-track” track is created and stored. If the track fit fails, or the collection of hits does not pass the basic quality requirements (for example because of a lack of hits) then the particle is said to be un-reconstructable. In this way, pseudo-tracking performance represents the ideal reconstruction performance given the ATLAS detector, with perfect hit-to-track association and track reconstruction efficiency. The approach was introduced in [57] as a way to obtain a fast approximation of tracking reconstruction for simulated data, however the technique has become a useful tool for studying tracking performance in general [42].

The ambiguity solver is not run for pseudo-tracks. However, if the standard track collection is produced alongside the pseudo-tracks, then cluster splitting neural networks will be run for the standard tracks, and the resulting classification of clusters will be propagated to hits on pseudo-tracks. This quirk allows one to study the inefficiencies of the cluster splitting process, and relatedly to determine whether shared hit cuts in the ambiguity solver are too loose or too tight. The fraction of hits that are shared for the IBL and the B-layer is shown in Fig. 3.6. The shared hits on pseudo-tracks represent correctly assigned hits from merged clusters that were not able to be classified as split by the cluster splitting neural networks. As such, these represent the number of shared hits the ambiguity solver should aim to allow. For shared hits on the IBL for particles produced before the IBL, the cuts appear to be successful in disallowing excessive numbers of shared hits. However, the ambiguity solver fails to limit shared hits for those particles produced after the IBL, reflecting the previously discussed problem of displaced tracks picking up incorrect hits. Meanwhile, it is clear that for the B-layer, the ambiguity solver is being overly aggressive in its rejection of shared hits.

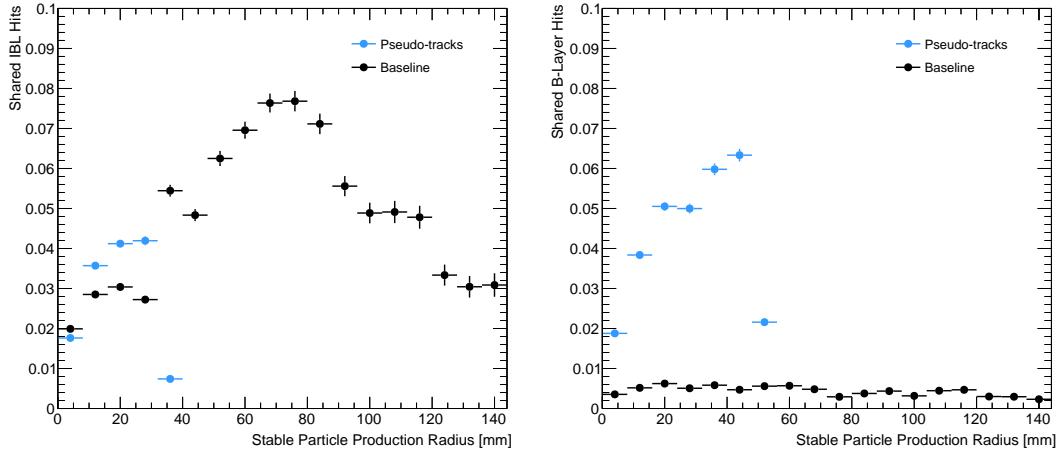


Figure 3.6: The rate of shared hits on b -hadron decay tracks on the IBL (left), and the B-layer (right), as a function of the production radius of the b -hadron decay product. Pseudo-tracks represent ideal performance given the ATLAS detector, and given also the efficiency of the NN cluster splitting algorithms.

3.2.2 Global χ^2 Fitter Outlier Removal

This section documents ongoing progress into improving hit assignments using the Global χ^2 Fitter (GX2F) to prevent wrong hits from being assigned to tracks during the track fit. This is in contrast to the approach discussed in [cref sec:refit](#), which attempts to identify and remove wrong hits after the reconstruction of the track (of which the track fit is a part). As part of the track fit, an outlier removal procedure is run, in which suspicious hits are identified and removed. The GX2F code, as a relatively low-level component of track reconstruction, has not undergone significant modification for several years. During this time, a new tracking sub-detector, the IBL, was installed, and subsequently precise detector alignments have been derived. The motivation for looking at the GX2F is that these changes may require re-optimisation of the GX2F code, and in particular the outlier removal procedures. Further motivation for this approach comes from the low rate of labelled outliers in baseline tracking. For example, while approximately 15% of B hadron decay tracks have a wrong IBL hit (a value which only increases with the p_T of the B), less than 1% of this tracks have had their IBL hit labelled and removed as an outlier. This section documents an attempt to improve hit assignment the Global χ^2 Fitter (GX2F) to prevent wrong hits from being assigned to tracks during the track fit. This is in contrast to the approach discussed in [crefsec:refit](#), which attempts to

discuss whether to keep this section

807 identify and remove wrong hits after the reconstruction of the track (of which the
 808 track fit is a part). As part of the track fit, an outlier removal procedure is run, in
 809 which suspicious hits are identified and removed. The motivation for this approach
 810 comes from the low rate of labelled outliers in baseline tracking. For example, while
 811 approximately 15% of B hadron decay tracks have a wrong IBL hit (a value which
 812 only increases with the p_T of the B), less than 1% of these tracks have had their IBL
 813 hit labelled and removed as an outlier.

814 Implementation

815 The outlier removal procedure for the pixel detector is described in this section. The
 816 states (also called measurements, or hits) on the track are looped over in order of
 817 increasing radial distance to the beam pipe. For each state, errors $\sigma(m_i)$ on the
 818 measurement of the transverse and longitudinal coordinates are calculated. These
 819 errors are dependent on the sub-detector which recorded the measurement (as some
 820 sub-detectors are more precise than others). Additionally, a residual displacement
 821 $r_i = m_i - x_i$ between the predicted position of the track x_i (inclusive of the current
 822 measurement), and the position of the measurement itself, m_i , is calculated. The
 823 pull p_i on the track state due to the current measurement is calculated according to

$$p_i = \frac{m_i - x_i}{\sqrt{\sigma(m_i)^2 - \sigma(x_i)^2}} \quad (3.2)$$

824 This pull is computed for the transverse and longitudinal coordinates of the mea-
 825 surement, and the maximum of the two is selected and checked to see if it exceeds a
 826 certain threshold. If it does, the hit will be removed, after some additional checks are
 827 made to confirm or deny the presence of the outlier. The threshold is set as a member
 828 variable `m_outlcut`. The results of varying this cut are described in Section 3.2.2.

829 Cut Optimisation

830 A systematic variation of the cut point `m_outlcut` has been carried out. The value of
 831 `m_outlcut` was reduced from 4 down to 1.75, a change which affects all silicon layers
 832 (the TRT has separate outlier removal logic). Furthermore, a specific cut for the IBL
 833 was introduced, and is set to 1.25. A second cut, `TrackChi2PerNDFCut`, is also used

834 in the outlier removal. This value was reduced from 7 to 4. Finally, instead of taking
 835 the maximum of the pulls in the longitudinal and transverse directions, a quadrature
 836 sum is taken of these two values and used. This variation is labelled “Mod GX2F” in
 837 plots.

838 The results, demonstrating a reduction in wrong hit assignment whilst also improving
 839 slightly the good good hits assigned to tracks, are shown in Fig. 3.7. The improve-
 840 ments are also observed when looking inclusively in all tracks, which removes the
 841 need for a specific *b*-jet ROI (a requirement which led to problems outlined in ??).
 An improvement, though modest, of all track parameter resolutions and pulls is

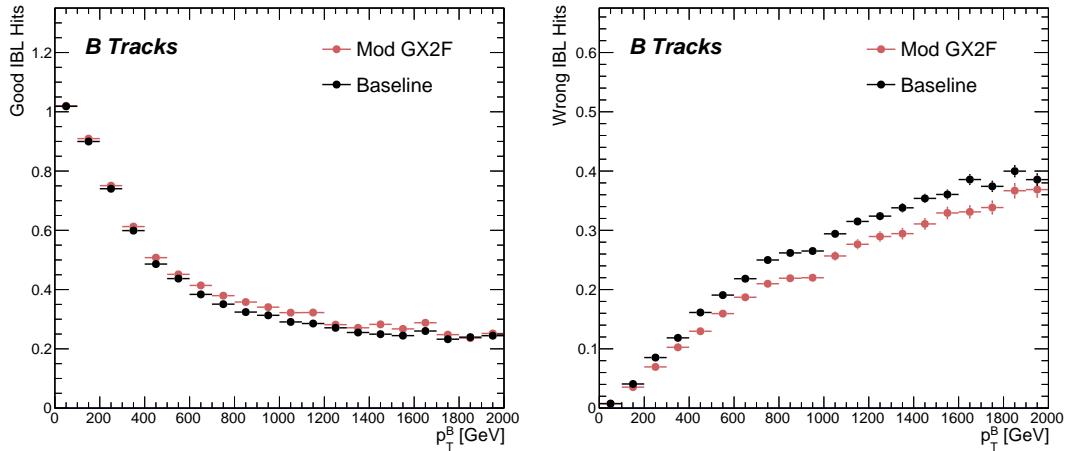


Figure 3.7: Profiles, as a function of parent *b*-hadron p_T , of good (left) and wrong (right) hit assignment rates on the IBL for tracks using baseline tracking (black), the modified version of the outlier removal procedure (red).

842
 843 observed. Some results are shown in Figs. 3.7 and 3.8, whilst the remainder of the
 844 plots can be found in the talks linked on the task’s Jira page. The results demon-
 845 strate an improvement in hit assignment, unchanged reconstruction efficiency, and
 846 modest improvement in track parameter resolutions and pulls. In addition, the truth
 847 match probability of track is unchanged, suggesting that there is no increase in fake
 848 track rates. The changes are expected to have a negligible impact on computational
 849 resources.

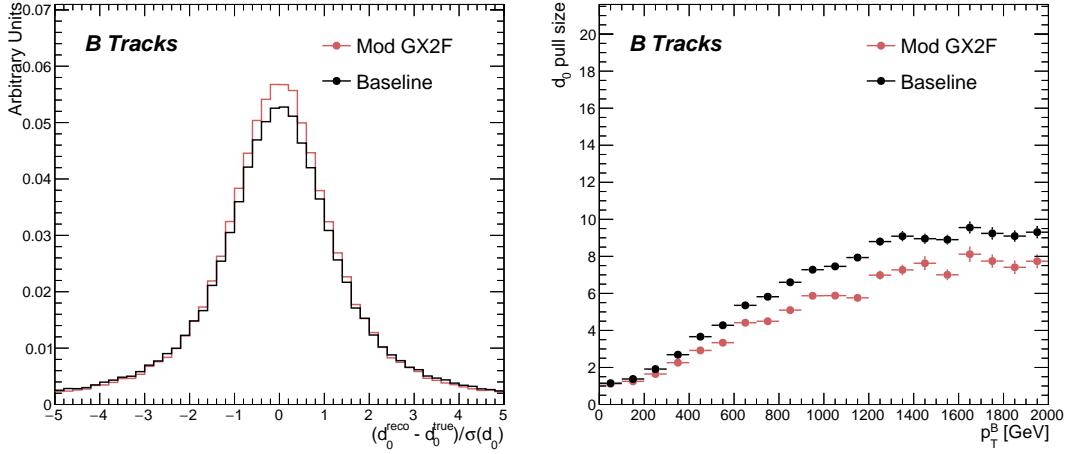


Figure 3.8: (left) B hadron decay track d_0 pulls for baseline and modified GX2F tracks. (right) The magnitude of the decay track d_0 pull as a function of B hadron transverse momentum.

850 3.3 Conclusion

851 In this section, the difficulties facing efficient and accurate track reconstruction,
 852 and hence performant b -tagging, have been outlined. The ambiguity solver, which
 853 attempts to clean or reject tracks which have excessive number shared hits, is shown
 854 to be overly aggressive in the removal of b -hadron decay product track candidates.
 855 The ambiguity solving process relies on many pre-defined cuts which have not been
 856 optimised for high transverse momentum b -hadron track reconstruction. These
 857 conclusions have motivated further ongoing studies into the improvement of the
 858 track reconstruction in dense environments and the high- p_T regime, such as those in
 859 Ref. [56].

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

869 **Chapter 4**

870 **Track Classification MVA**

871 The chapter details work on implementing a multivariate algorithm (MVA) to predict
872 the truth origin of reconstructed tracks. An introduction to formalisms of machine
873 learning is given in Section 4.1. In Section 4.2, the truth origin label is defined,
874 and in Section 4.3 these labels are used to train a machine learning model that can
875 effectively discriminate between good and fake tracks. Several studies motivated this
876 work by demonstrating that at high p_T , b -tagging performance was degraded by the
877 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm
878 could be trained to detect fake tracks, these could be removed before their input to
879 the b -tagging algorithms with the aim of improving performance.

880 **4.1 Machine Learning Background**

881 Over the past few decades, machine learning (ML) techniques have become in-
882 creasingly popular in high energy physics experiments due the increased volumes
883 of high-dimensional data and improvements in the field of machine learning (in
884 particular deep learning). Machine learning is the process in which a computer
885 program uses data to learn suitable parameters for a predictive model model. This
886 is opposed to explicitly providing instructions on how to perform a task. A subfield
887 known as *supervised learning* is used in this work, and consists of exposing a model
888 to a large number of labelled examples in order to extract relationships between the
889 input data and their labels. These relationships are often complex, and explicitly

890 programmed rules can fail to fully capture the relationships between inputs and
 891 outputs.

892 In the simplest case, a set of m labelled training examples $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$
 893 is collected. Each element (x_i, y_i) consists of a input vector $x_i \in \mathbb{R}^{\text{input}}$, and the
 894 corresponding label y_i . In classification problems, these labels are integer *class*
 895 *labels* $y_i \in \{0, \dots, N - 1\}$, where N is the number of classes, which specify which
 896 of a pre-determined set of categorical classes the training example belongs to. The
 897 rest of the discussion in this chapter is limited to binary classification problems
 898 ($N = 2$). The two classes are often referred to as signal ($y_i = 1$) and background
 899 ($y_i = 0$), which need to be separated. Collecting sufficient and suitable data is one
 900 of the primary challenges of machine learning, as such data is not always readily
 901 available. Fortunately, sophisticated tools to simulate particle collisions have already
 902 been developed by the scientific community [58, 59]. These tools play a key role in
 903 generating a suitablly large amount of labelled data which is used to train algorithms.
 904 More detail on the input datasets is given in Section 4.3.1.

905 After obtaining suitable training data, the next step is to define a model. Given an
 906 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
 907 parameterised functional mapping from input space to output space. Given an input
 908 example x_i and a set of parameters θ , the model outputs a prediction $\hat{y}_i \in (0, 1)$ for
 909 the true label y_i , as in

$$f_\theta(x_i) = \hat{y}_i. \quad (4.1)$$

910 The output \hat{y}_i is in the interval $(0, 1)$ so as to be interpreted as the probability
 911 that the input example x_i belongs to the signal class. The parameters θ of the
 912 model are randomly initialised, however the model is designed to be expressive
 913 enough to correctly map the inputs x_i to the outputs y_i given the correct choice of
 914 parameters. The model is then trained, which amounts to showing the model a series
 915 of labelled training examples and modifying the parameters of the model based on
 916 its performance.

917 4.1.1 Neural Networks

918 Neural networks (NNs) are a common choice for the machine learning model f since
 919 they have the ability to approximate any function [60] and are easy to train via
 920 backpropagation [61].

921 Artificial Neurons

922 The basic functional component of a NN is the *artificial neuron* or node, which is
 923 loosely inspired by a mathematical model of a biological neuron [62, 63]. An artificial
 924 neuron is defined by its parameters or *weights* θ and a choice of activation function.
 925 Each neuron takes a fixed number of inputs and computes the dot product of the
 926 input and weight vectors $x^T \theta$ and additionally adds a constant bias term θ_0 . This
 927 term plays the role of a trainable constant value that is independent of the inputs.

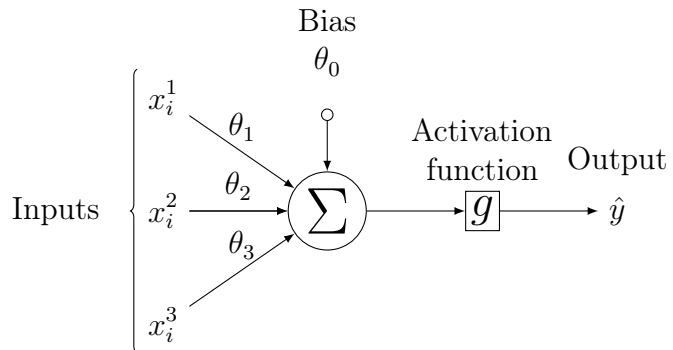


Figure 4.1: A diagram displaying the logical flow of a single neuron with three inputs x_i^j . Each input is multiplied by a weight θ_j , and the resulting values are summed. A bias term θ_0 is added, and the result is passed to an activation function. Each neuron can be thought of as a logistic regression model.

928 The dot product is fed into an activation function g . The activation function has
 929 several uses, most notably acting as a source of non-linearity and bounding the
 930 output of the neuron. Some common activation functions are shown in Fig. 4.2.
 931 The choice of activation function can have implications for the performance and
 932 convergence of the network, since the gradient of g is used to compute the weight
 933 updates during training. This is also why input data is normalised to have zero mean
 934 and unity variance [64].

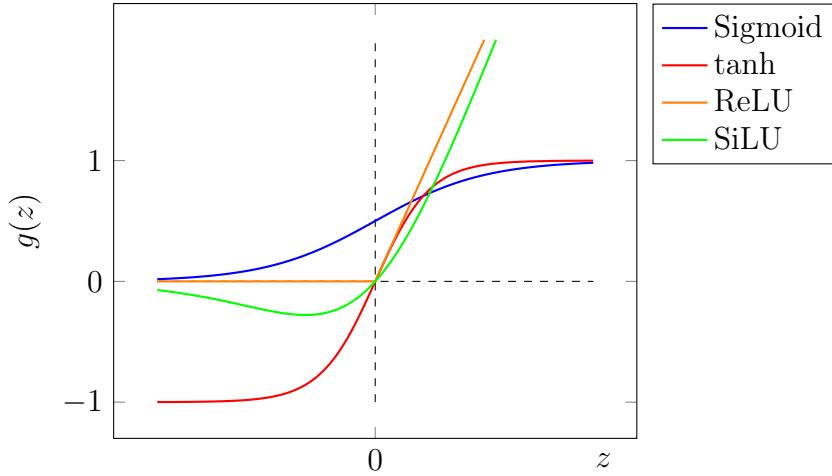


Figure 4.2: Several common choices for the activation function g of an artificial neuron.

935 Networks

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

$$g(z) = \frac{1}{1 + e^{-z}}, \quad (4.2)$$

942 which is bounded between zero and one allowing the final output to be interpreted
 943 as the probability that the input sample belongs to the signal class. NNs have the
 944 crucial property of being differentiable functions, which facilitates training process
 945 described in the next section.

946 4.1.2 Training with Gradient Descent

947 A training algorithm is used to optimise the weights of a NN after exposure to the
 948 training data. The training algorithm works minimising a loss function L , which
 949 quantifies the error in the model's predictions for a given input. NNs are commonly
 950 trained using backpropagation in combination with a variant of stochastic gradient

descent to iteratively update the model parameters. In binary classification problems, the binary cross entropy given in Eq. (4.3) loss if often used.

$$L(x_i, \theta) = y_i \ln[f_\theta(x_i)] + (1 - y_i) \ln[1 - f_\theta(x_i)] \quad (4.3)$$

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

$$\theta'_i = \theta_i - \alpha \frac{\partial L}{\partial \theta_i} \quad (4.4)$$

The hyperparameter α is known as the *learning rate* and dictates the size of the step taken in the direction of the slope. The errors for each parameter are efficiently calculated using the backpropagation algorithm. The process of updating weights is repeated until the weights converge the network is trained. In practice, small batches of the input data are shown to the network at a time. For each batch the average loss is calculated and the network's weights are updated. There are many extensions and variations of the gradient descent algorithm. This work uses the Adam optimiser which adds momentum to the weight updates (dampening oscillations) and an adaptive per-parameter learning rate [65].

4.2 Track Truth Origin Labelling

Crucial to supervised learning techniques are the ground truth class labels which the machine learning model is trained to predict. A set of track truth labels which a high degree of granularity have been implemented in the ATLAS software stack, and are listed in Table 4.1. The labelling scheme has designed to be useful beyond the classification of good and fake tracks. The origins are determined by analysing the detailed simulated truth record for the truth particle associated with each track. Tracks are associated with truth particles by selecting the truth particle with the highest *truth-matching probability* (TMP), defined in Eq. (4.5). This is a weighted sum of the number of hits on a track which are from the same truth particle, versus the total number of hits on the track. The weights are subdetector-dependent are

977 designed to account for the varying number of layers in each of the subdetectors.

$$\text{TMP} = \frac{10N_{\text{Pix}}^{\text{good}} + 5N_{\text{SCT}}^{\text{good}} + N_{\text{TRT}}^{\text{good}}}{10N_{\text{Pix}}^{\text{all}} + 5N_{\text{SCT}}^{\text{all}} + N_{\text{TRT}}^{\text{all}}} \quad (4.5)$$

978 For the fake track classification tool, the origins in Table 4.1 are used to construct a
 979 binary label by labelling all fake tracks background, and all other tracks as signal.
 980 The fake track classifier is then trained to distinguish between these two categories
 981 of tracks. Fake tracks are also defined using the TMP, with a $\text{TMP} < 0.5$ giving a
 982 track the label of fake. Fake tracks are made up of combinatorial fakes, which are
 983 tracks which do not correspond to the trajectory of any truth particle, and poorly
 984 reconstructed tracks, which may somewhat resemble the trajectory of a truth particle
 985 but may be off due to the presence of some wrong hits on the track.

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 4.1: 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 [43]. 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.

986 4.3 Fake Track Identification Tool

987 The rate of fake tracks increases at high transverse momentum as shown in Fig. 4.3 due
 988 to the difficulties in track reconstruction outlined in Section 3.1.2. The performance
 989 of b -tagging algorithms is reduced as a direct result of the presence of these fake
 990 tracks as shown for SV1 in Fig. 4.4.

991 To identify and remove fake tracks, a NN classification tool was trained with good
 992 tracks as the signal class and fake tracks as the background class. Due to the
 993 imbalance between the two classes (with fake tracks being relatively uncommon), a
 994 weight was added to the loss function for the background class to account for this.
 995 The NN was made up of two hidden layers with 100 nodes per layer. The ReLU
 996 activation function was used in conjunction with the Adam optimiser with a learning
 997 rate of $1e-3$. Optimisation of the networks architecture has been carried out to
 998 ensure good performance without a relatively small number of learnable parameters.
 999 The model was trained using 40 million tracks with a further 1 million tracks each
 1000 used for validation and testing.

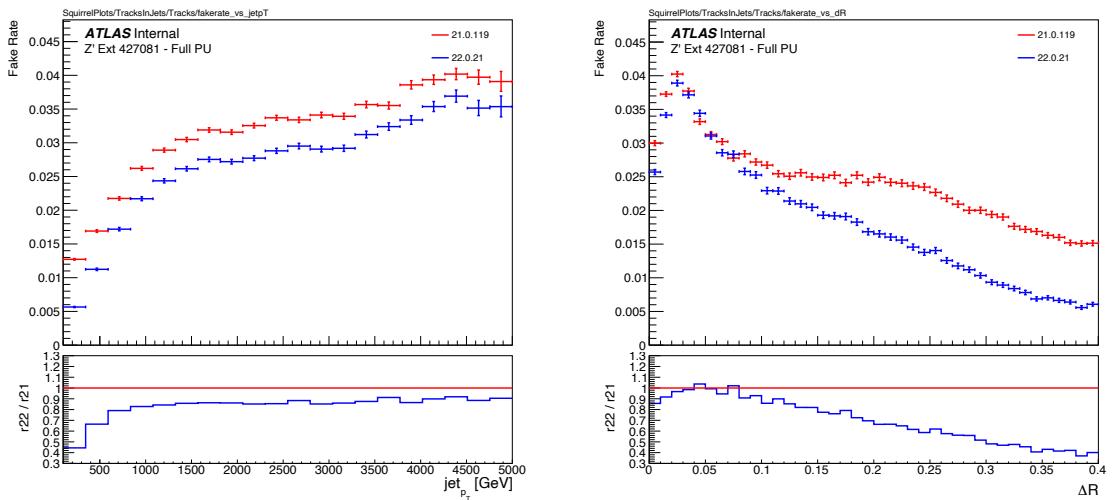


Figure 4.3: Rate of fake tracks as a function of jet transverse momentum (left) and ΔR (track, jet) (right). The rate of fake tracks increases significantly as a function of p_T , and also increases as the distance to the jet axis decreases, and the number of tracks in the jet increases (not shown).

1001 Inputs to the model are described in Section 4.3.2, while fake track removal perfor-
 1002 mance is given in Section 4.3.3.

1003 4.3.1 Datasets

1004 To train and evaluate the model, simulated SM $t\bar{t}$ and BSM Z' events initiated by
 1005 proton-proton collisions at a center of mass energy $\sqrt{s} = 13$ TeV are used. The Z'
 1006 sample is constructed in such a manner that it has a relatively flat jet p_T spectrum

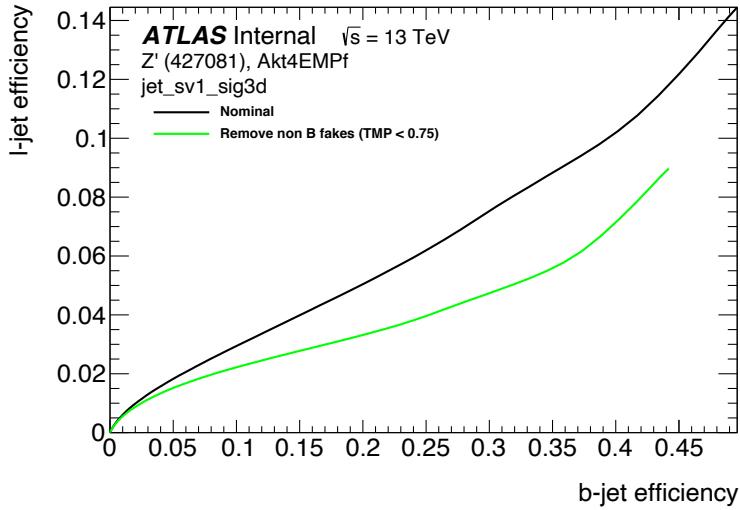


Figure 4.4: The light-jet efficiency of the low level tagger SV1 as a function of b -jet efficiency for the nominal tracking setup (black) and for the case where fake tracks which are not from the decay of a b -hadron are removed. The light-jet efficiency is decreased, demonstrating that the presence of fake tracks is detrimental to algorithm performance.

1007 up to 5 TeV and decays to an equal numbers of b -, c - and light- jets. The generation
 1008 of the simulated event samples includes the effect of multiple pp interactions per
 1009 bunch crossing with an average pileup of $\langle \mu \rangle = 40$, which includes the effect on the
 1010 detector response due to interactions from bunch crossings before or after the one
 1011 containing the hard interaction.

1012 The $t\bar{t}$ events are generated using the POWHEGBOX [66–69] v2 generator at next-
 1013 to-leading order with the NNPDF3.0NLO [70] set of parton distribution functions
 1014 (PDFs). The h_{damp} parameter¹ is set to 1.5 times the mass of the top-quark (m_{top}) [71],
 1015 with $m_{\text{top}} = 172.5$ GeV. The events are interfaced to PYTHIA 8.230 [72] to model the
 1016 parton shower, hadronisation, and underlying event, with parameters set according
 1017 to the A14 tune [73] and using the NNPDF2.3LO set of PDFs [74]. Z' events are
 1018 generated with PYTHIA 8.2.12 with the same tune and PDF set. The decays of b -
 1019 and c -hadrons are performed by EVTGEN v1.6.0 [75]. Particles are passed through
 1020 the ATLAS detector simulation [76] based on GEANT4 [77].

¹The 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.

4.3.2 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 4.2. 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 track divided by the total number of tracks in the event. The ambiguity solver processes track candidates iteratively (see Section 2.3.1), and the order in which tracks are accepted is preserved. Tracks which do not require cleaning (i.e. the removal of suspect shared hits) are likely to be accepted earlier on. Hence the FracRank variable gives an indication of how easy it was for the track to be reconstructed.

Track selection follows the loose selection described in Ref. [78] and outlined in Table 4.3, which was found to improve the flavour tagging performance compared to previous tighter selections, whilst ensuring good resolution of tracks and a low fake rate [43]. Inputs are scaled to have a central value of zero and a variance of unity before training and evaluation.

4.3.3 Results

Performance of the fake track classification tool was run on 1 million tracks in jets in the combined $t\bar{t}$ and Z' samples. The continuous scalar output from the NN model is interpreted as the probability that a given track is a good track (i.e. not fake). Fig. 4.5 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 in 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. Considering only tracks from b -hadron decays, this value drops slightly to 0.928.

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 4.2: 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 [43], while split hits are hits used on multiple tracks which have been identified as merged.

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 4.3: 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. Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [43]. Shared hits on pixel layers are given a weight of 1, while shared hits in the SCT are given a weight of 0.5. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

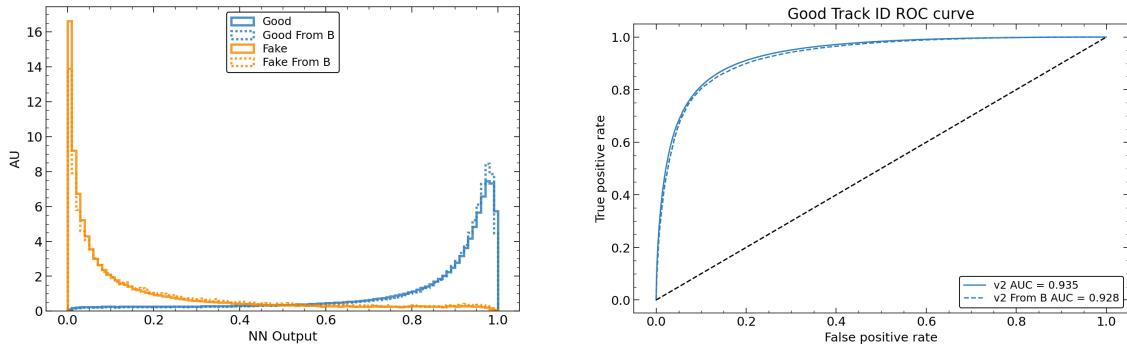


Figure 4.5: (left) Normalised histogram of the model output separated for good and fake tracks, and further separated by those tracks which are from the decay of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a b -hadron (dashed line).

1049 Good and fake track efficiencies at two different NN output cut points are shown
1050 in Table 4.4. The results demonstrate that the tool is effective in retaining 98.8%
1051 of good tracks, while correctly identifying (and therefore enabling the removal of)
1052 45.6% of fake tracks. Table 4.4 also shows that a significant amount of tracks which
1053 are labelled as both fake and from the decay of a b -hadron are also removed.

MVA Output Cut	Good 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 4.4: Good and fake track selection efficiencies for the combined $t\bar{t}$ and Z' samples. Two working points are defined, cutting on the NN output at 0.06 and 0.12. The continuous output of the model allows for the tuning of good and fake track identification efficiencies.

1054 After initial tests and investigation, it was found that fake tracks which were the
1055 result of b -hadron decays actually aided b -tagging performance. The application of
1056 a single tool which removed all fake tracks was no longer optimal. A second tool
1057 was trained in the same manner of the first, this one was designed to distinguish
1058 between those tracks which were from the decay of a b -hadron and those which were
1059 not. A 2-dimensional cut was then used to only reject those tracks that had a high
1060 probability of being fake, and also a low probability of being a b -hadron decay track.
1061 The light-jet efficiency of SV1 shows an improvement when using the combined tools
1062 to remove fake tracks that are not from a b -hadron decay, as shown in Fig. 4.6

also show
plots for
JF?

4.4 Conclusion

1064 Fake tracks, which are prevalent in the core of high p_T jets, are shown to have an
1065 adverse impact on b -tagging performance. A ML tool to identify fake tracks has
1066 been developed, which can be used to limit the number of fake tracks being inputted
1067 to downstream b -tagging algorithms. It was found that, since many b -hadron decay
1068 tracks are poorly reconstructed and thus marked as fake, it was necessary also to
1069 train a second algorithm to detect b -hadron decay tracks so that the removal of
1070 these tracks could be avoided. Removing fake and non- b decay tracks in this way

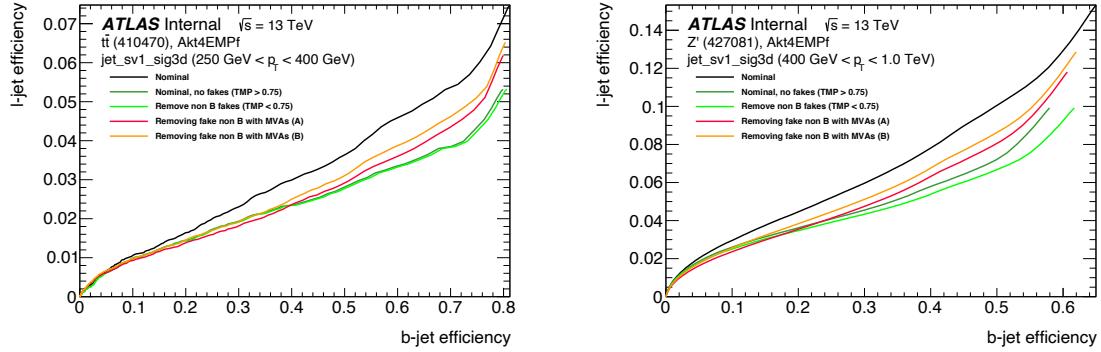


Figure 4.6: The effect of applying the fake track identification algorithm alongside the b -hadron decay track identification on the jet tagging performance of SV1 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black) is compared to two working points of fake track removal (orange and red). Removal of fake tracks based on truth information is shown by the green curves.

1072 was found to improve the tagging efficiency of SV1 and JetFitter at high transverse
1073 momentum.

1074 While removing tracks prior to their input to the low level tagging algorithms is
1075 beneficial, a more performant alternative would be to keep these tracks but label
1076 them as being fake or poorly reconstructed, and allow the tagging algorithms to take
1077 this into consideration, potentially still making use of some of the information.

1078 These tools, which identify the origin of a given track, have other potential uses.
1079 One application is to isolate a relatively pure sample of fake tracks which can be
1080 used to estimate the fake track rate in data, which would be useful for producing
1081 the recommendations for tracking systematic uncertainties. Another application
1082 would be to use the b -hadron track identification tool to improve the track-to-jet
1083 association.

1084 The approach here works on a track-by-track basis, but a more sophisticated approach
1085 would consider the correlations between the tracks inside a jet, as shown in Chapter 5.

1086 Also left for future work is to simultaneously train a single tool which discriminates
1087 between all the truth origins listed in Table 4.1. Such a tool would be useful as a
1088 general purpose multiclass classifier.

₁₀₈₉ **Chapter 5**

₁₀₉₀ **Graph Neural Network Flavour
1091 Tagger**

₁₀₉₂ Some of the work in this chapter has previously been published in Ref. [79]. The
₁₀₉₃ author of this thesis was on the editorial team. Figures, tables and text from the
₁₀₉₄ published note are reproduced here.

₁₀₉₅ Flavour tagging, the identification of jets originating from b - and c -quarks, is a
₁₀₉₆ critical component of the physics programme of the ATLAS experiment at the Large
₁₀₉₇ Hadron Collider. Current flavour tagging algorithms rely on the outputs of several
₁₀₉₈ low-level algorithms, which reconstruct various properties of jets using charged
₁₀₉₉ particle tracks, that are then combined using machine learning techniques. In this
₁₁₀₀ note a new machine learning algorithm based on graph neural networks, GN1, is
₁₁₀₁ introduced. GN1 uses information from a variable number of charged particle tracks
₁₁₀₂ within a jet, to predict the jet flavour without the need for intermediate low-level
₁₁₀₃ algorithms. Alongside the jet flavour prediction, the model predicts which physics
₁₁₀₄ processes produced the different tracks in the jet, and groups tracks in the jet into
₁₁₀₅ vertices. These auxiliary training objectives provide useful additional information on
₁₁₀₆ the contents of the jet and improve performance. GN1 compares favourably with
₁₁₀₇ the current ATLAS flavour tagging algorithms. For a b -jet efficiency of 70%, the
₁₁₀₈ light (c)-jet rejection is improved by a factor of ~ 1.8 (~ 2.1) for jets coming from
₁₁₀₉ $t\bar{t}$ decays with transverse momentum $20 < p_T < 250$ GeV. For jets coming from Z'
₁₁₁₀ decays with transverse momentum $250 < p_T < 5000$ GeV, the light (c)-jet rejection
₁₁₁₁ improves by a factor ~ 6 (~ 2.8) for a comparative 30% b -jet efficiency.

5.1 Motivation

Flavour tagging, the identification of jets originating from b - and c -quarks, is a critical component of the physics programme of the ATLAS experiment [30] at the Large Hadron Collider (LHC) [80]. It is of particular importance for the study of the Standard Model (SM) Higgs boson and the top quark, which preferentially decay to b -quarks [81, 82], and additionally for several Beyond Standard Model (BSM) resonances that readily decay to heavy flavour quarks [83]. The significant lifetime of b -hadrons, approximately 1.5 ps [84], provides the unique signature of a secondary decay vertex which has a high mass and is significantly displaced from the primary vertex. Additional signatures of b -hadrons are the tertiary decay vertex, resulting from $b \rightarrow c$ decay chains, and the reconstructed trajectories of charged particles (henceforth simply referred to as tracks) with large impact parameters¹ (IPs). These signatures are primarily identified using tracks associated to jets. As such, efficient and accurate track reconstruction is essential for high performance flavour tagging.

This note introduces a novel algorithm, GN1, which uses Graph Neural Networks (GNNs) [85] with auxiliary training objectives, to aid the primary goal of classifying whether jets originate from b - or c -quarks (referred to as a flavour tagger). The concept is illustrated in Fig. 5.1. The use of GNNs offers a natural way to classify jets with variable numbers of unordered associated tracks, while allowing for the inclusion of auxiliary training objectives [86, 87].

The current ATLAS flavour tagger, DL1r [88], is a deep neural network which takes the outputs of a number of independently optimised “low-level” algorithms [47] as inputs. Each of these low-level algorithms makes use of tracks to reconstruct a particular aspect of the experimental signature of heavy flavour jets. The low-level algorithms can be manually optimised reconstruction algorithms, for example the SV1 and JetFitter algorithms that reconstruct displaced decay vertices, or trained taggers such as RNNIP and DIPS that use the IPs of a variable number of tracks to identify the flavour of the jet [47, 78, 89, 90]. In contrast GN1 utilises a single neural network, which directly takes the tracks and some information about the jet as inputs. As such, it does not depend on any other flavour tagging algorithm, and a single training of the GN1 fully optimises all aspects of the algorithm.

¹The distance of closest approach from a track to the primary vertex.

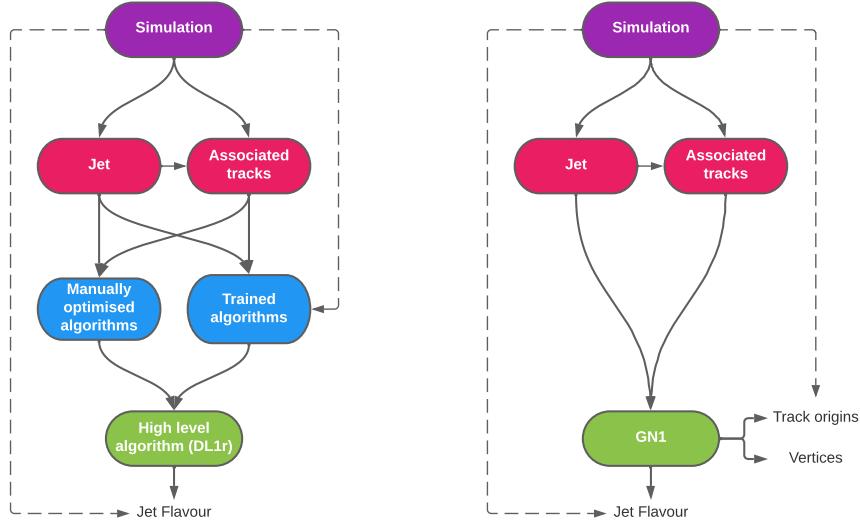


Figure 5.1: Comparison of the existing flavour tagging scheme (left) and GN1 (right). The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, additional 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.

1143 GN1 is trained to understand the internal structure of the jet through the use of
1144 two auxiliary training objectives: the grouping of tracks originating from a common
1145 vertex, and the prediction of the underlying physics process from which each track
1146 originated. These auxiliary objectives are meant to guide the neural network towards
1147 a more complete understanding of the underlying physics, removing the need for the
1148 low-level algorithms, and therefore simplifying the process of optimising the tagger for
1149 new regions of phase space (e.g. c -tagging or high- p_T b -tagging), or when the detector
1150 or charged particle reconstruction algorithms are updated. The training targets
1151 for the primary and auxiliary objectives are extracted from “truth information”, i.e.
1152 information only available in simulation, as opposed to reconstructed quantities
1153 available in both collision data and simulation.

1154 In this note, the following benefits of this approach will be shown:

- 1155 1. Improved performance with respect to the current ATLAS flavour tagging
1156 algorithms, with larger background rejection for a given signal efficiency.
 - 1157 2. The same network architecture can be easily optimised for a wider variety of use
1158 cases (e.g. c -jet tagging and high- p_T jet tagging), since there are no low-level
1159 algorithms to retune.
 - 1160 3. There are fewer flavour tagging algorithms to maintain.
 - 1161 4. Alongside the network’s prediction of the jet flavour, the auxiliary vertex and
1162 track origin predictions provide more information on why a jet was (mis)tagged
1163 or not. This information can also have uses in other applications, for instance
1164 to explicitly reconstruct displaced decay vertices or to remove fake tracks.²
- 1165 This note is organised as follows: a brief description of the ATLAS detector, object
1166 definitions and selections, and samples are provided in Section 5.3; details about the
1167 model architecture and training procedure are given in Section 5.4; and results are
1168 discussed in Section 5.5.

²A 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. [43].

₁₁₆₉ **5.2 Graph Neural Network Theory**

₁₁₇₀ **5.3 Experimental Setup**

₁₁₇₁ **5.3.1 Datasets**

₁₁₇₂ Datasets used to train the GN1 tagger are the same as described in Section 4.3.1.

₁₁₇₃ Jets are also required not to overlap with a generator-level electron or muon from
₁₁₇₄ W boson decays. All jets are required to have a pseudorapidity $|\eta| < 2.5$ and
₁₁₇₅ $p_T > 20 \text{ GeV}$. Additionally, a standard selection using the Jet Vertex Tagger (JVT)
₁₁₇₆ algorithm at the tight working point is applied to jets with $p_T < 60 \text{ GeV}$ and $|\eta| < 2.4$
₁₁₇₇ in order to suppress pile-up contamination [91].

₁₁₇₈ As previously, truth labelled b -, c - and light- jets are kinematically re-sampled in
₁₁₇₉ p_T and η to ensure identical distributions in these variables. The resulting dataset
₁₁₈₀ contains 30 million jets, 60% of which are $t\bar{t}$ jets and 40% of which are Z' jets. While
₁₁₈₁ DL1r uses 70% $t\bar{t}$ jets and 30% Z' jets, the change in sample composition did not
₁₁₈₂ affect the final performance of GN1. To evaluate the performance of the model, 500k
₁₁₈₃ jets from both the $t\bar{t}$ and Z' samples, which are statistically independent from the
₁₁₈₄ training sample, are used. Track- and jet-level inputs are scaled to have a central
₁₁₈₅ value of zero and a variance of unity before training and evaluation.

₁₁₈₆ **5.4 Model Architecture**

₁₁₈₇ **5.4.1 Model Inputs**

₁₁₈₈ GN1 is given two jet variables and 21 tracking related variables for each track
₁₁₈₉ fed into the network. The jet transverse momentum and signed pseudorapidity
₁₁₉₀ constitute the jet-level inputs, with the track-level inputs listed in Table 5.1. If a jet
₁₁₉₁ has more than 40 associated tracks, the first 40 tracks with the largest transverse

1192 IP significance³ $s(d_0)$ are selected as inputs. Full track parameter information
1193 and associated uncertainties, along with detailed hit information, carry valuable
1194 information about the jet flavour. In the dense cores of high- p_T jets, tracks are highly
1195 collimated and separation between tracks can be of the same order as the active
1196 sensor dimensions, resulting in merged clusters and tracks which share hits [43]. Due
1197 to the relatively long lifetimes of b -hadrons and c -hadrons, which can traverse several
1198 layers of the ID before decaying and have highly collimated decay products, the
1199 presence of shared or missing hits is a critical signature of heavy flavour jets. Full
1200 track parameter information and associated uncertainties, along with detailed hit
1201 information, carry valuable information about the jet flavour.

1202 Dependence on the absolute value of the azimuthal jet angle ϕ is explicitly removed
1203 by providing only the azimuthal angle of tracks relative to the jet axis. The track
1204 pseudorapidity is also provided relative to the jet axis.

1205 Since heavy flavour hadrons can decay semileptonically, the presence of a recon-
1206 structed lepton in the jet carries discriminating information about the jet flavour.
1207 In addition to the baseline GN1 model, the GN1 Lep variant includes an additional
1208 track-level input, leptonID, which indicates if the track was used in the reconstruction
1209 of an electron, a muon or neither. The muons are required to be combined [92],
1210 and the electrons are required to pass the *VeryLoose* likelihood-based identification
1211 working point [93].

1212 Track selection follows the loose selection used for the fake track classification MVA
1213 in Chapter 4 and described in Table 4.3. This selection was found to improve the
1214 flavour tagging performance compared to previous tighter selections, whilst ensuring
1215 good resolution of tracks and a low fake rate [43]. Section 5.6.1 demonstrates that
1216 further relaxation of the track selection requirements may be warranted.

1217 5.4.2 Auxiliary Training Objectives

1218 In addition to the jet flavour classification, two auxiliary training objectives are
1219 defined. Each auxiliary training objective comes with a training target which, similar

3Impact 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 [34].

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 5.1: 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 [43], 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 GN1 Lep model.

1220 to the jet flavour label, are truth labels derived from the simulation. The presence
1221 of the auxiliary training objectives improves the jet classification performance as
1222 demonstrated in Section 5.5.4.

1223 The first auxiliary objective is the prediction of the origin of each track within the
1224 jet. Each track is labelled with one of the exclusive categories defined in Table 4.1
1225 after analysing the particle interaction that led to its formation. Since the presence
1226 of different track origins is strongly related to the flavour of the jet, training GN1 to
1227 recognise the origin of the tracks may provide an additional handle on the classification
1228 of the jet flavour. This task may also aid the jet flavour prediction by acting as a
1229 form of supervised attention [94] - in detecting tracks from heavy flavour decays the
1230 model may learn to pay more attention to these tracks.

1231 Displaced decays of b - and c -hadrons lead to secondary and tertiary vertices inside
1232 the jet. Displaced secondary vertices can also occur in light-jets as a result of material
1233 interactions and long-lived particle decays (e.g. K_S^0 and Λ^0). The second auxiliary
1234 objective is the prediction of track-pair vertex compatibility. For each pair of tracks
1235 in the jet, GN1 predicts a binary label, which is given a value 1 if the two tracks
1236 in the pair originated from the same point in space, and 0 otherwise. To derive the
1237 corresponding truth labels for training, truth production vertices within 0.1 mm are
1238 merged. Track-pairs where one or both of the tracks in the pair have an origin label
1239 of either Pileup or Fake are given a label of 0. Using the pairwise predictions from
1240 the model, collections of commonly compatible tracks can be grouped into vertices.
1241 The addition of this auxiliary training objective removes the need for inputs from a
1242 dedicated secondary vertexing algorithm.

1243 Both auxiliary training objectives can be considered as “stepping stones” on the way
1244 to classifying the flavour of the jet. By requiring the model to predict the truth
1245 origin of each track and the vertex compatibility of each track-pair, the model is
1246 guided to learn representations of the jet which are connected to the underlying
1247 physics and therefore relevant for classifying the jet flavour.

1248 5.4.3 Architecture

1249 As discussed above, the GN1 model combines a graph neural network architecture [95]
1250 with auxiliary training objectives in order to determine the jet flavour. Coarse

optimisation of the network architecture hyperparameters, for example number of layers and number of neurons per layer, has been carried out to maximise the tagging efficiency.

The model architecture is based on a previous implementation of a graph neural network jet tagger [87]. As compared to the previous approach, GN1 uses a only a single graph neural network and makes use of a more sophisticated graph neural network layer [96], described below. These changes yield improved tagging performance and a significant reduction in training time with respect to the previous approach.

The model takes jet- and track-level information as inputs, as detailed in Section 5.4.1. The jet inputs are concatenated with each track’s inputs, as shown in Fig. 5.2. The combined jet-track vectors are then fed into a per-track initialisation network with three hidden layers, each containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 5.3. The track initialisation network is similar to a Deep Sets model [97], but does not include a reduction operation (mean or summation) over the output track representations.

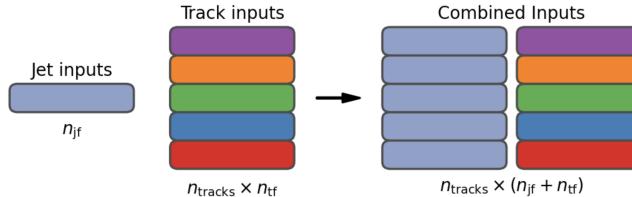


Figure 5.2: The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of n_{tracks} , where each track is described by 21 track features ($n_{\text{tf}} = 21$). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1.

A fully connected graph is built from the outputs of the track initialisation network, such that each node in the graph neighbours every other node. Each node h_i in the graph corresponds to a single track in the jet, and is characterised by a feature vector, or representation. The per-track output representations from the initialisation networks are used to populate the initial feature vectors of each node in the graph. In each layer of the graph network, output node representations h'_i are computed by aggregating the features of h_i and neighbouring nodes \mathcal{N}_i as described in Ref. [96]. First, the feature vectors of each node are fed into a fully connected layer \mathbf{W} , to produce an updated representation of each node $\mathbf{W}h_i$. These updated feature vectors are used to compute edge scores $e(h_i, h_j)$ for each node pair,

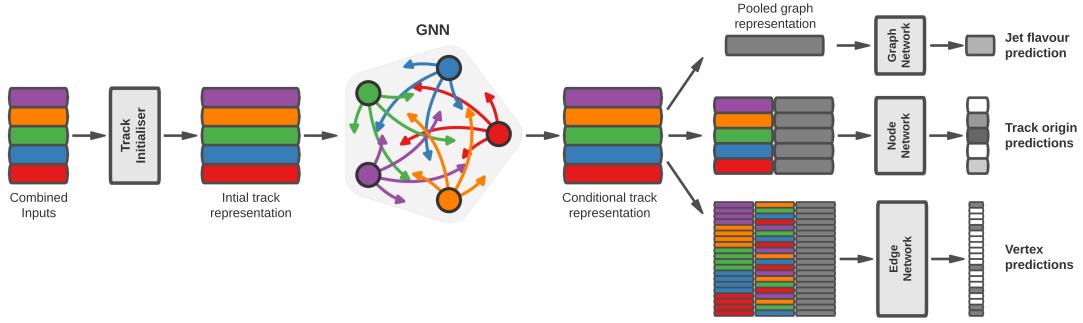


Figure 5.3: The network architecture of GN1. Inputs are fed into a per-track initialisation network, which outputs an initial latent representation of each track. These representations are then used to populate the node features of a fully connected graph network. After the graph network, the resulting node representations are used to predict the jet flavour, the track origins, and the track-pair vertex compatibility.

$$e(h_i, h_j) = \mathbf{a}^\perp \theta [\mathbf{W}h_i \oplus \mathbf{W}h_j], \quad (5.1)$$

1276 where \oplus denotes vector concatenation, θ is a non-linear activation function, and \mathbf{a} is
 1277 a learned vector. These edge scores are then used to calculate attention weights a_{ij}
 1278 for each pair of nodes using the softmax function over the edge scores

$$a_{ij} = \text{softmax}_j [e(h_i, h_j)]. \quad (5.2)$$

1279 Finally, the updated node representation h'_i is computed by taking the weighted sum
 1280 over each updated node representation $\mathbf{W}h_i$, with weights a_{ij}

$$h'_i = \sigma \left[\sum_{j \in \mathcal{N}_i} a_{ij} \cdot \mathbf{W}h_j \right]. \quad (5.3)$$

1281 The above set of operations constitute a single graph network layer. Three such layers
 1282 are stacked to construct the graph network, representing a balance between achieving

optimal performance and preventing overtraining. The final output node feature vectors from the network are representations of each track that are conditional on the other tracks in the jet. The output representation for each track is combined using a weighted sum to construct a global representation of the jet, where the attention weights for the sum are learned during training. Three separate fully connected feedforward neural networks are then used to independently perform the different classification objectives of GN1. Each of the objectives makes use of the global representation of the jet. A summary of the different classification networks used for the various training objectives is shown in Table 5.2.

Table 5.2: 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

A node classification network, which takes as inputs the features from a single output node from the graph network and the global jet representation, predicts the track truth origin, as defined in Table 4.1. This network has three hidden layers containing 128, 64 and 32 neurons respectively, and an output size of seven, corresponding to the seven different truth origins.

An edge classification network, which takes as inputs the concatenated representations from each pair of tracks and the global jet representation, is used to predict whether the tracks in the track-pair belong to a common vertex. The edge network has three hidden layers containing 128, 64 and 32 neurons respectively, and a single output, which is used to perform binary classification of the track-pair compatibility. These predictions are used for the auxiliary training objectives discussed in Section 5.4.2.

A graph classification network takes only the global jet representation as an input, and predicts the jet flavour. The graph classification network is comprised of four fully connected hidden layers with 128, 64, 32 and 16 neurons respectively, and has three outputs corresponding to the b -, c - and light- jet classes.

1307 5.4.4 Training

1308 The full GN1 training procedure minimises the total loss function L_{total} , defined in
 1309 Eq. (5.4). This loss is composed of three terms: L_{jet} , the categorical cross entropy
 1310 loss over the different jet flavours; L_{vertex} , the binary track-pair compatibility cross
 1311 entropy loss averaged over all track-pairs; and L_{track} , the categorical cross entropy loss
 1312 for the track origin prediction. L_{vertex} is computed by averaging over all track-pairs
 1313 in the batch, and L_{track} is computed by averaging over all tracks in the batch.

$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}} \quad (5.4)$$

1314 The different losses converge to different values during training, reflective of differences
 1315 in the relative difficulty of the various objectives. As such, L_{vertex} and L_{track} are
 1316 weighted by $\alpha = 1.5$ and $\beta = 0.5$ respectively to ensure they converge to similar values,
 1317 giving them an equal weighting towards L_{total} . The values of α and β also ensure
 1318 that L_{jet} converges to a larger value than L_{vertex} and L_{track} , reflecting the primary
 1319 importance of the jet classification objective. In practice, the final performance of
 1320 the model was not sensitive to modest variations in the loss weights α and β , or to
 1321 pre-training using L_{total} and fine tuning on the jet classification task only. As there
 1322 was a significant variation in the relative frequency of tracks of different origins, the
 1323 contribution of each origin class to L_{track} was weighted by the inverse of the frequency
 1324 of their occurrence. In L_{vertex} , the relative class weight in the loss for track-pairs
 1325 where both tracks are from either a b - or c -hadron is increased by a factor of two as
 1326 compared with other track-pairs.

1327 The track classification and vertexing objectives are supplementary to the jet clas-
 1328 sification objective and trainings can be performed with either the node or edge
 1329 networks, or both, removed, as discussed in Section 5.5.4. In these cases, the cor-
 1330 responding losses L_{vertex} and L_{track} are removed from the calculation of L_{total} . The
 1331 resulting trainings demonstrate how useful the different auxiliary training objectives
 1332 are for the primary jet classification objective.

1333 GN1 trainings are run for 100 epochs on 4 NVIDIA V100 GPUs, taking around 25
 1334 mins to complete each epoch over the training sample of 30 million jets described
 1335 in Section 5.3.1. The Adam optimiser [98] 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, stabilised after around 60 epochs. The epoch that minimized the validation loss was used for evaluation. GN1 has been integrated into the ATLAS software [40] using ONNX [99], and jet flavour predictions for the test sample are computed using the ATLAS software stack.

5.5 Results

The performance of the GN1 tagger is evaluated for both b -tagging and c -tagging use cases, and for both jets with $20 < p_T < 250$ GeV from the $t\bar{t}$ sample and jets with $250 < p_T < 5000$ GeV from the Z' sample. Performance is compared to the DL1r tagger [88], which has been retrained on 75 million jets from the same samples as GN1. The input RNNIP tagger [90] to DL1r has not been retrained.

The taggers predict 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 , defined as

$$D_b = \log \frac{p_b}{(1 - f_c)p_l + f_c p_c}, \quad (5.5)$$

where f_c is a free parameter that determines the relative weight of p_c to p_l in the score D_b , controlling the trade-off between c - and light-jet rejection performance. This parameter is set to a value of $f_c = 0.018$ for the DL1r model, obtained through an optimisation procedure designed to maximise the c - and light-jet rejection of DL1r [88]. For the GN1 models a value of $f_c = 0.05$ is used, based on a similar optimisation procedure. The choice of f_c is arbitrary, with the different optimised values reflecting the relative c - versus light-jet rejection performance of the various taggers. A fixed-cut working point (WP) defines the corresponding selection applied to the tagging discriminant D_b in order to achieve a given inclusive efficiency on the $t\bar{t}$ sample.

The technical implementation of GN1 results in any jet with no associated tracks or exactly one associated track to be classified as a light-jet. The impact of this on the

tagging performance of GN1 was found to be negligible, with 0.12% of b -jets in the $t\bar{t}$ sample and 0.02% of b -jets in the Z' sample affected. Of those, 89% of the b -jets in the $t\bar{t}$ sample and 98% of the b -jets in the Z' sample are classified as light-jets by DL1r at the 70% $t\bar{t}$ WP.

A comparison of the b -tagging discriminant D_b between DL1r and GN1 is given in Fig. 5.4. The shapes of the distributions are broadly similar for b -, c - and light-jets, however, the GN1 model shifts the b -jet distribution to higher values of D_b in the regions with the best discrimination. The GN1 c -jet distribution is also shifted to lower values of D_b when compared with DL1r, enhancing the separation and indicating that GN1 will improve c -jet rejection when compared with DL1r.

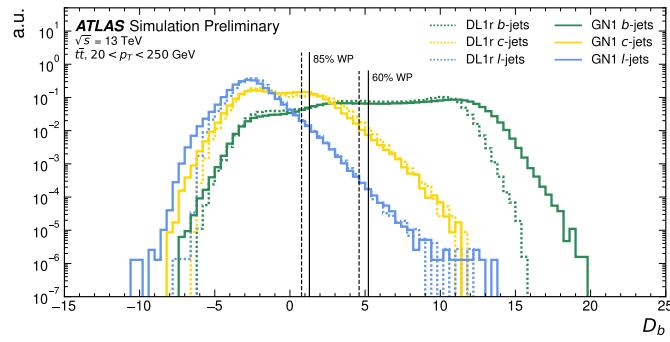


Figure 5.4: Comparison between the DL1r and GN1 b -tagging discriminant D_b for jets in the $t\bar{t}$ sample. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs used by analyses. A value of $f_c = 0.018$ is used in 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.

5.5.1 b -tagging Performance

The performance of a b -tagging algorithm is quantified by its power to reject c - and light-jets for a given b -jet tagging efficiency, or WP. In order to compare the b -tagging performance of the different taggers for the b -jet tagging efficiencies in the range typically used by analyses, the corresponding c - and light-jet rejection rates are displayed in Figs. 5.5 and 5.6 for jets in the $t\bar{t}$ and Z' samples respectively. Four standard WPs with b -jet tagging efficiencies of 60%, 70%, 77% and 85% are used by physics analyses depending on their specific signal and background requirements. These WPs are defined using jets in the $t\bar{t}$ sample only. The b -jet tagging efficiencies

for jets in the Z' sample are lower than the corresponding WPs calculated in the $t\bar{t}$ sample, due to the much higher jet p_T range in the Z' sample. For instance the WP defined to provide a 70% b -jet tagging efficiency on the $t\bar{t}$ sample results in a b -jet tagging efficiency of $\sim 30\%$ on the Z' sample. To account for this, the range of b -jet tagging efficiencies displayed in Fig. 5.6 is chosen to span the lower values achieved in the Z' sample.

For jets in the $t\bar{t}$ sample with $20 < p_T < 250 \text{ GeV}$, GN1 demonstrates considerably better c - and light-jet rejection compared with DL1r across the full range of b -jet tagging efficiencies probed. The relative improvement depends on the b -jet tagging efficiency, with the largest improvements found at lower values. At a b -jet tagging efficiency of 70%, the c -rejection improves by a factor of ~ 2.1 and 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 $250 < p_T < 5000 \text{ GeV}$, GN1 also brings considerable performance improvements with respect to DL1r across the range of b -jet tagging efficiencies studied. Again, the largest relative improvement in performance comes at lower b -jet tagging efficiencies. At a b -jet tagging efficiency of 30%, GN1 improves the c -rejection by a factor of ~ 2.8 and the light-jet rejection by a factor of ~ 6 . An increasing statistical uncertainty due to the high rejection of background affects the comparison at lower b -jet tagging efficiencies. 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 the sophisticated exploitation of low-level information, further studies are needed to confirm if the performance gain is also observed in experimental data.

The GN1 Lep variant shows improved performance with respect to the baseline GN1 model, demonstrating 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 GN1 Lep. The improvement in light-jet rejection also increases from a factor of ~ 1.8 to ~ 2.5 at this 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. 5.7, the greatest improvement of GN1 Lep over GN1 is seen at low p_T .

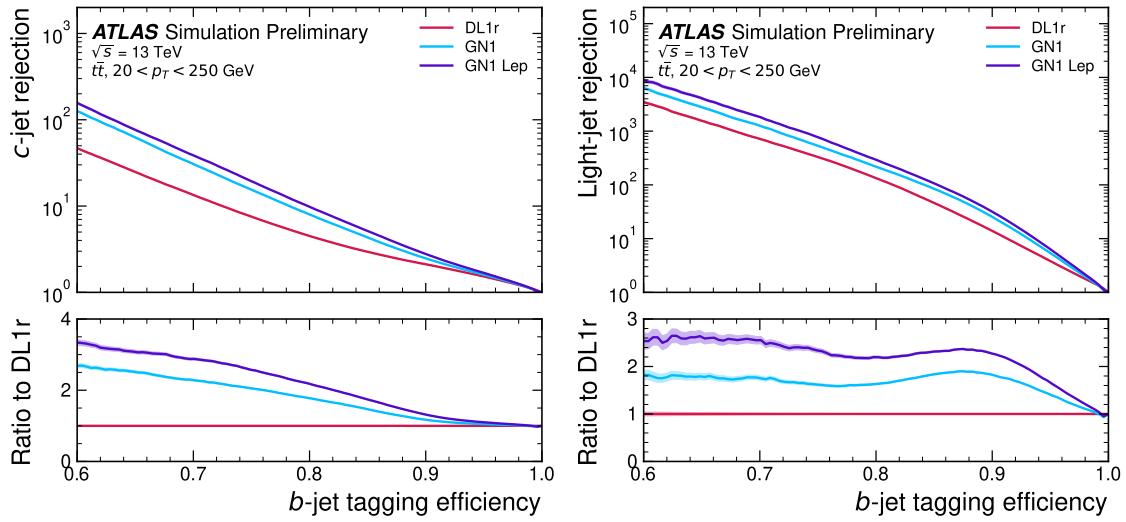


Figure 5.5: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. 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.

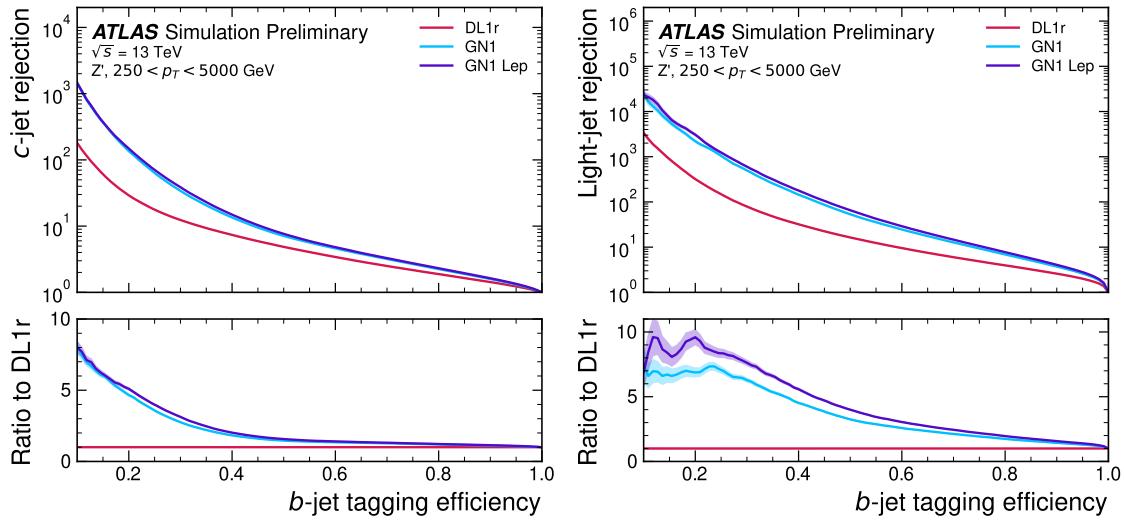


Figure 5.6: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000 \text{ GeV}$. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. 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.

1415 The performance of the taggers is strongly dependent on the jet p_T . Charged particle
1416 reconstruction is particularly challenging within high- p_T jets [43]. The multiplicity of
1417 fragmentation particles increases as a function of p_T , while the number of particles
1418 from heavy flavour decays stays constant. Collimation of particles inside the jet
1419 increases and approaches the granularity of the tracking detectors, making it difficult
1420 to resolve the trajectories of different particles. Furthermore, at high p_T , heavy
1421 flavour hadrons will travel further into the detector before decaying. For hadrons
1422 which traverse one or more layers of the ID before decaying, the corresponding decay
1423 tracks may pick up incorrect hits, left by the hadron itself or fragmentation particles,
1424 in the inner layers of the detector, reducing the accuracy of the reconstructed track
1425 parameters. These factors contribute to a reduced reconstruction efficiency for heavy
1426 flavour tracks, and a general degradation in quality of tracks inside the core of a jet,
1427 which in turn reduces the jet classification performance.

1428 In order to study how the b -jet tagging efficiency of the taggers varies as a function
1429 of jet p_T , the b -jet tagging efficiency as a function of p_T for a fixed light-jet rejection
1430 of 100 in each bin is shown in Fig. 5.7. For jets in the $t\bar{t}$ sample, at a fixed light-jet
1431 rejection of 100, GN1 improves the b -jet tagging efficiency by approximately 4%
1432 across all jet p_T bins. GN1 Lep shows improved performance with respect to GN1,
1433 in particular at lower p_T , with the relative increase in the b -jet tagging efficiency
1434 going from 4% to 8%. For jets in the Z' sample, GN1 has a higher b -jet tagging
1435 efficiency than DL1r across the p_T range, with the largest relative improvement in
1436 performance, approximately a factor of 2, found at jet $p_T > 2$ TeV. GN1 outperforms
1437 DL1r across the entire jet p_T spectrum studied. The performance was also evaluated
1438 as a function of the average number of pileup interactions in an event, and was found
1439 to have no significant dependence on this quantity.

1440 5.5.2 c -tagging Performance

1441 Since GN1 does not rely on any manually optimised low-level tagging algorithms,
1442 which may not have been optimised for c -tagging, tagging c -jets presents a compelling
1443 use case for GN1. To use the model for c -tagging, the output probabilities are
1444 combined into a single score D_c , defined similarly to Eq. (5.5) as

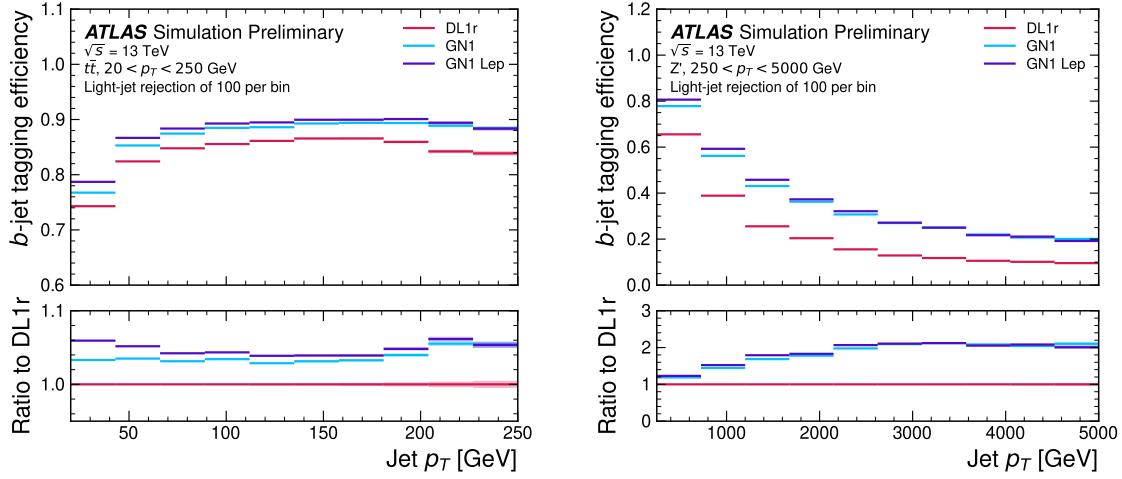


Figure 5.7: The b -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T with a fixed light-jet rejection of 100 in each bin. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands are denoted by the shaded regions.

$$D_c = \log \frac{p_c}{(1 - f_b)p_l + f_b p_b}. \quad (5.6)$$

A value of $f_b = 0.2$ is used for all models. Similar to Section 5.5.1, performance of the different taggers is compared by scanning through a range of c -jet tagging efficiencies and plotting the corresponding b - and light-jet rejection rates. As in Section 5.5.1, WPs are defined using jets in the $t\bar{t}$ sample. Standard c -jet tagging efficiency WPs are significantly lower in comparison with the b -tagging WPs in order to maintain reasonable b - and light-jet rejection rates. This is reflected in the range of c -jet tagging efficiencies used in Figs. 5.8 and 5.9. In Fig. 5.8, which displays the c -tagging performance of the models on the jets in the $t\bar{t}$ sample, GN1 performs significantly better than DL1r. The b - and light-jet rejection improve most at lower c -jet tagging efficiencies, with both background rejections increasing by a factor of 2 with respect to DL1r at a c -jet tagging efficiency of 25%. GN1 Lep outperforms GN1, with the b -rejection (light-jet rejection) relative improvement increasing from a factor of 2 to 2.1 (2 to 2.3) at the 25% c -jet WP. Fig. 5.9 shows the c -tagging performance on the jets in the Z' sample. Both GN1 and GN1 Lep perform similarly,

1459 improving the b -rejection by 60% and the light-jet rejection by a factor of 2 at the
1460 25% c -jet WP.

1461 **5.5.3 Jet Display Diagrams**

1462 **5.5.4 Ablations**

1463 Several ablations, the removal of components in the model to study their impact,
1464 are carried out to determine the importance of the auxiliary training objectives of
1465 GN1 to the overall performance. The “GN1 No Aux” variant retains the primary jet
1466 classification objective, but removes both track classification and vertexing auxiliary
1467 objectives (see Section 5.4.2) and as such only minimises the jet classification loss.
1468 The “GN1 TC” variant includes track classification but not vertexing, while “GN1
1469 Vert” includes vertexing, but not track classification.

1470 For jets in both the $t\bar{t}$ and Z' samples, the models without one or both of the auxiliary
1471 objectives display significantly reduced c - and light-jet rejection when compared with
1472 the baseline GN1 model, as shown in Figs. 5.12 and 5.13. For jets in the $t\bar{t}$ sample,
1473 the performance of GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert
1474 perform similarly to each other. For jets in the Z' sample, the GN1 No Aux model
1475 shows a clear improvement in c - and light-jet rejection when compared with DL1r at
1476 lower b -jet tagging efficiencies. Similar to jets in the $t\bar{t}$ sample, GN1 TC and GN1
1477 Vert perform similarly, and bring large gains in background rejection when compared
1478 with GN1 No Aux, but the combination of both auxiliary objectives yields the best
1479 performance.

1480 It is notable that the GN1 No Aux model matches or exceeds the performance of
1481 DL1r without the need for inputs from the low-level algorithms. This indicates that
1482 the performance improvements enabled by GN1 appear to be able to compensate for
1483 the removal of the low-level algorithm inputs. The GN1 TC and GN1 Vert variants
1484 each similarly outperform DL1r, demonstrating that both contribute to the overall
1485 high performance of the baseline model.

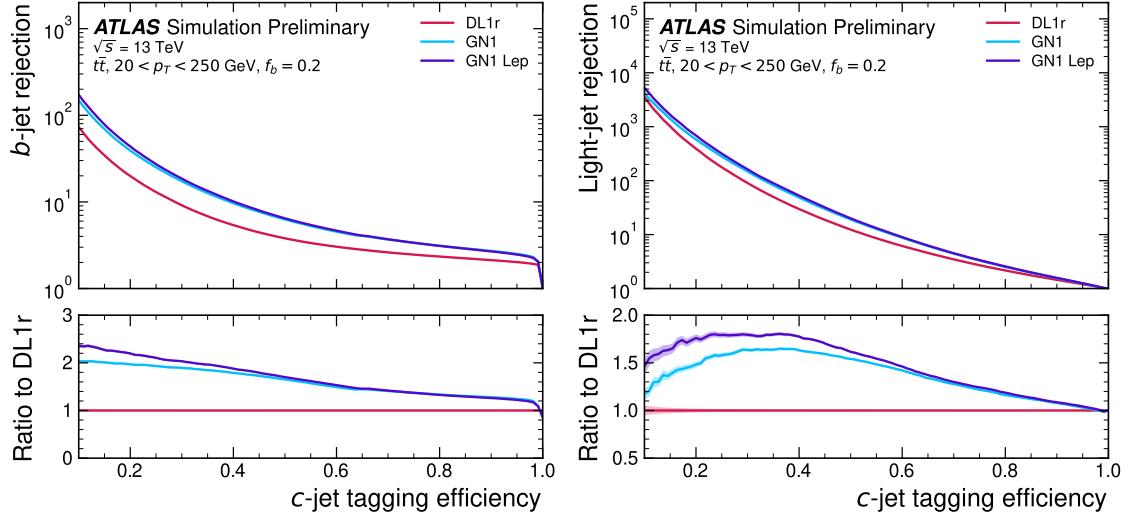


Figure 5.8: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. At c -jet 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.

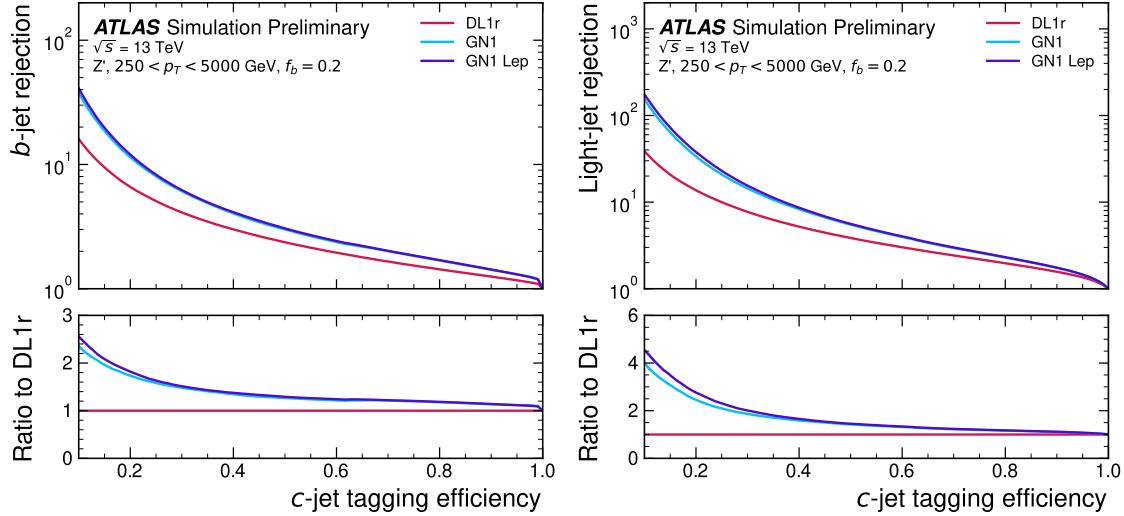


Figure 5.9: The b -jet (left) and light-jet (right) rejections as a function of the c -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The lower x -axis range is chosen to display the c -jet tagging efficiencies usually probed in these regions of phase space.

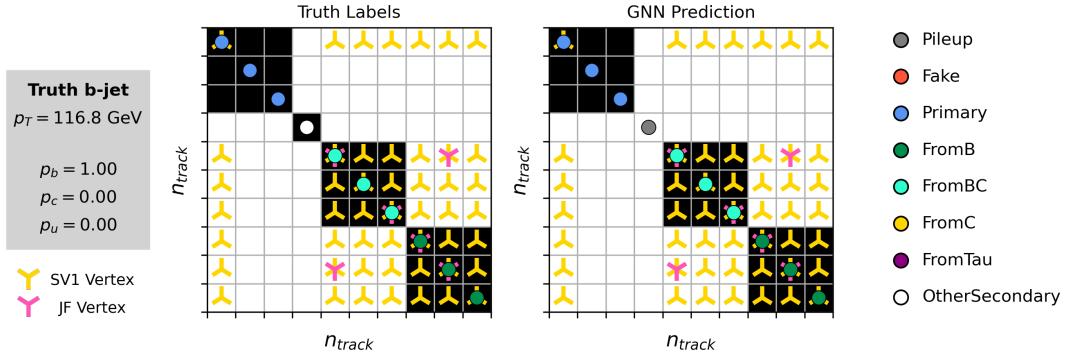


Figure 5.10: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked.

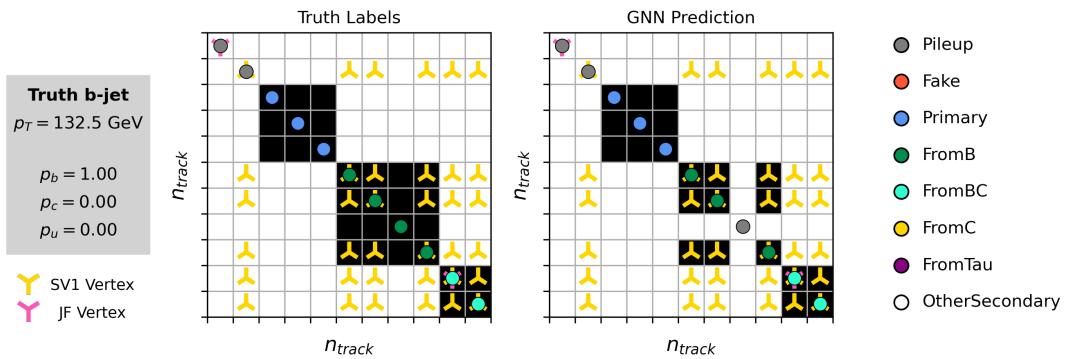


Figure 5.11: A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). Vertices reconstructed by SV1 and JetFitter are also marked.

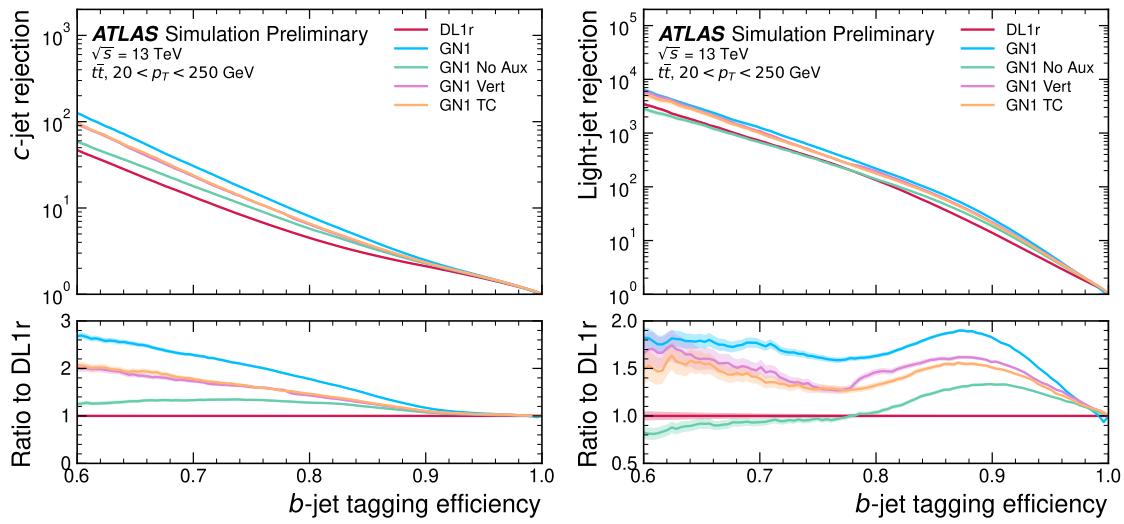


Figure 5.12: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (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.

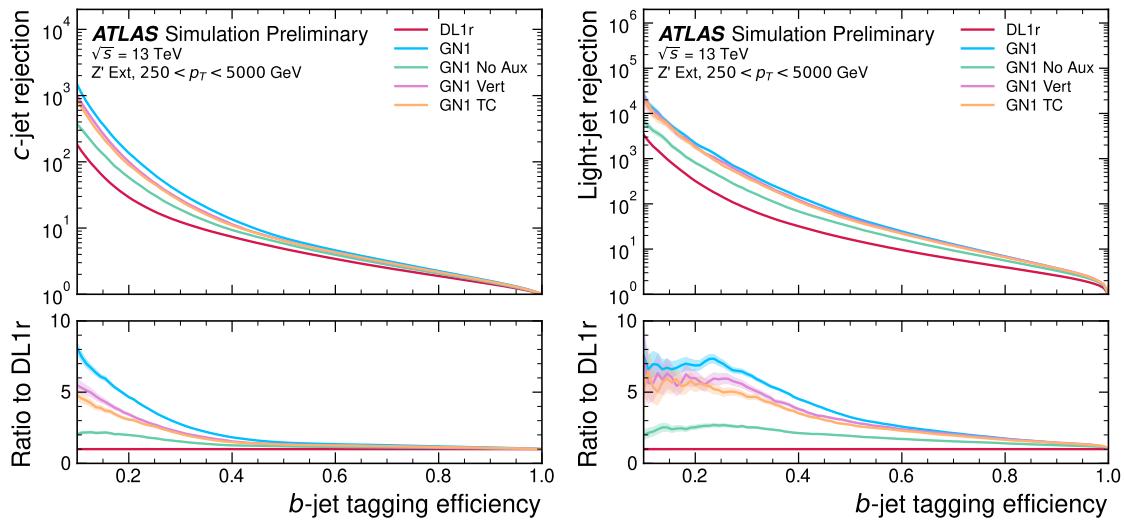


Figure 5.13: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary 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.

1486 5.5.5 Inclusion of Low-Level Vertexing Algorithms

1487 GN1 does not include inputs from low-level tagging algorithms, including the vertexing
1488 tools SV1 and JetFitter [47]. Since these algorithms are known to improve the
1489 performance of DL1r, it was feasible that their inclusion in GN1 may further improve
1490 on the performance of the GN1 models. In a dedicated training of GN1 the SV1 and
1491 JetFitter tagger outputs were added to the GN1 jet classification network as an input,
1492 similar to their use in DL1r. These outputs include information on the reconstructed
1493 vertices, including the number of vertices, the vertex mass, displacement, and other
1494 properties. In addition, the index of the reconstructed SV1 or JetFitter vertices were
1495 included as two track-level inputs to GN1. The jet classification performance of this
1496 GN1 model was not significantly different to the baseline model, and in some cases
1497 the performance was slightly reduced. A dedicated look at the vertexing performance
1498 of GN1 with some comparisons to SV1 and JetFitter is found in Section 5.5.6

1499 5.5.6 Vertexing Performance

1500 From the track-pair vertex prediction described in Section 5.4.2, tracks can be
1501 partitioned into compatible groups representing vertices (see [87]). As such, GN1
1502 is able to be used to perform vertex “finding”, but not vertex “fitting”, i.e. the
1503 reconstruction of a vertex’s properties, which currently still requires the use of a
1504 dedicated vertex fitter. In order to study the performance of the different vertexing
1505 tools inside b -jets, the truth vertex label of the tracks, discussed in Section 5.4.2, are
1506 used. To estimate the efficiency with which GN1 manages to find vertices inclusively,
1507 vertices from GN1 containing tracks identified as coming from a b -hadron are merged
1508 together and compared to the inclusive truth decay vertices that result from a
1509 b -hadron decay (where if there are multiple distinct truth vertices from a b -hadron
1510 decay they are also merged together). Vertices are compared with the target truth
1511 vertex and the number of correctly and incorrectly assigned tracks is computed.
1512 Since secondary vertex information is only recovered for reconstructed tracks, an
1513 efficiency of 100% here denotes that all possible secondary vertices are recovered
1514 given the limited track reconstruction efficiency. A vertex is considered matched if
1515 it contains at least 65% of the tracks in the corresponding truth vertex, and has a
1516 purity of at least 50%. GN1 manages to achieve an inclusive reconstruction efficiency

1517 in b -jets of $\sim 80\%$, demonstrating that it effectively manages to identify the displaced
1518 vertices from b -hadron decays.

1519 **More detail**

1520 In order to study the performance of the different vertexing tools inside b -jets, the
1521 truth vertex label of the tracks, discussed in Section 5.4.2, is used. The reconstructed
1522 vertices from GN1, SV1 and JetFitter are compared to the target truth vertices in
1523 order to calculate the efficiencies of the different vertexing tools. Since secondary
1524 vertex information is only recovered for reconstructed tracks, an efficiency of 100%
1525 here denotes that all possible secondary vertices are recovered given the limited track
1526 reconstruction efficiency.

1527 There are several caveats to a comparison of the vertexing tools which are a result
1528 of the different approaches they take to vertexing. SV1 and JetFitter are designed
1529 to only find secondary vertices in the jet, whereas GN1 is also trained to determine
1530 which tracks in the jet belong to the primary vertex (the vertex of the hard scatter
1531 pp interaction). To account for this the GN1 vertex with the largest number of
1532 predicted primary tracks is excluded from the vertex finding efficiency calculation.
1533 While JetFitter and GN1 aim to resolve each displaced vertex inside the jet, such
1534 that secondary vertices from b -hadron decays are found separately to tertiary vertices
1535 from $b \rightarrow c$ decay chains, SV1 by design attempts to find a single inclusive vertex
1536 per jet. This inclusive vertex groups inclusive b -hadron decays. These are tracks
1537 from the b -hadron decay itself (FromB) and tracks from $b \rightarrow c$ decays (FromBC).
1538 In order to fairly compare the performance if the different tools, both the exclusive
1539 and inclusive vertex finding efficiency is studied. For the exclusive vertex finding
1540 case JetFitter and GN1 can be directly compared, while a comparison with SV1 is
1541 not possible due to aforementioned design constraints. The inclusive vertex finding
1542 performance of all three tools can be compared using the procedure outlined below.

1543 The starting point for the secondary vertex finding efficiency in both the exclusive
1544 and inclusive cases is to select truth secondary vertices are those containing only
1545 inclusive b -hadron decays to be considered as initial targets. For exclusive vertex
1546 finding, these truth secondary vertices can be used directly as the denominator for the
1547 efficiency calculation. Meanwhile for the inclusive efficiency all such truth secondary
1548 vertices in the jet are merged into a single inclusive target vertex. Correspondingly,

for the inclusive vertex finding case, the vertices found by JetFitter are merged into a single vertex, and the vertices found by GN1 with at least one predicted inclusive b -hadron decay track are also merged similarly. SV1 does not require any vertex merging.

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

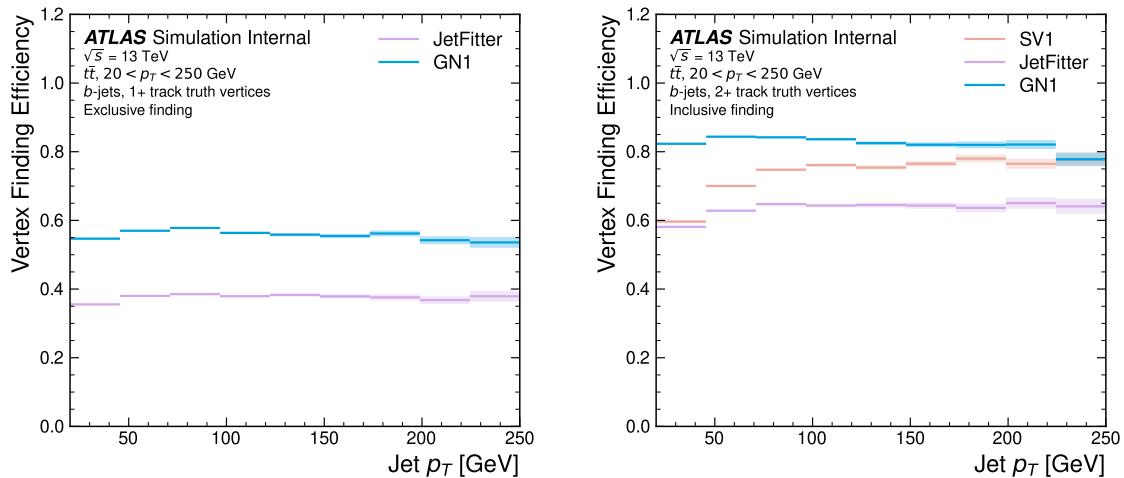


Figure 5.14: Vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% 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. 5.14. For b -jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV, the exclusive vertex finding efficiency of JetFitter and GN1 is relatively flat as a function of p_T . 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 $p_T > 100$ 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. 5.15 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.

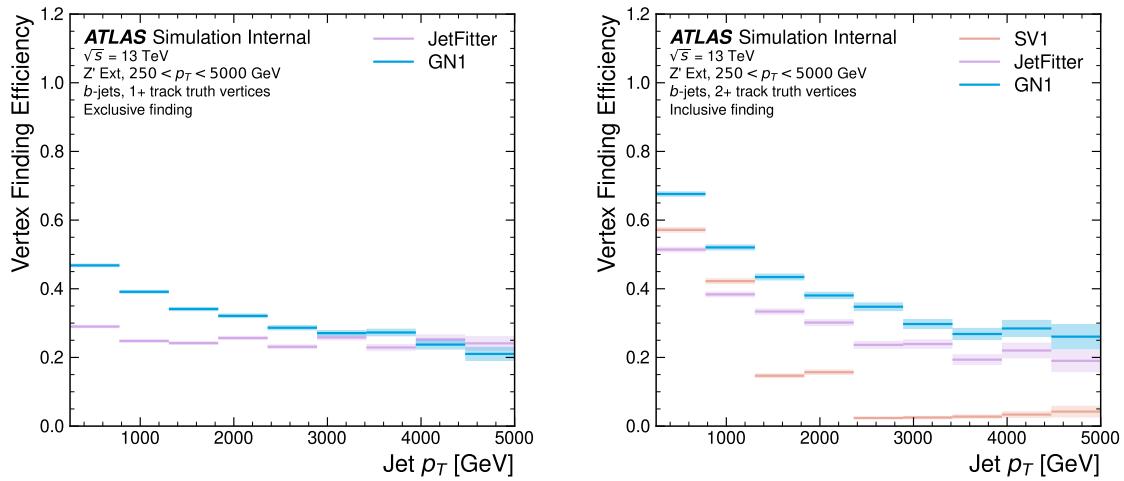


Figure 5.15: Inclusive vertex finding efficiency for multitrack truth vertices in b -jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T . Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly.

5.5.7 Track Classification Performance

As discussed in Section 5.4.2, one of the auxiliary training objectives for GN1 is to predict the truth origin of each track in the jet. Since the equivalent information is not provided by any of the existing flavour tagging tools, as a benchmark a multi-class classification multilayer perceptron (MLP) is trained on the same tracks used for the baseline GN1 training. The model uses the same concatenated track-and-jet inputs as GN1 (see Section 5.4.1), but processes only a single track at a time. The model is comprised of five densely connected layers with 200 neurons per layer, though the performance was not found to be strongly sensitive to changes in the network structure. To measure the track classification performance, the area under the curve (AUC) of the receiver operating characteristic (ROC) curve is computed for each origin class using a one versus all classification approach. The AUCs for the different truth origin classes are averaged using both an unweighted and a weighted approach. The unweighted mean treats the performance of each class equally, while the weighted mean uses the fraction of tracks from each origin as a weight. As seen in Table 5.3, GN1 outperforms the MLP, both at $20 < p_T < 250 \text{ GeV}$ for jets in the $t\bar{t}$ sample, and at $250 < p_T < 5000 \text{ GeV}$ for jets in the Z' sample. For tracks in jets in the $t\bar{t}$ sample, GN1 can reject 65% of fake tracks while retaining more than 99% of good tracks. The GN1 model has two advantages over the MLP which can explain the performance improvement. Firstly, the mixing of information between tracks, enabled by the fully connected graph network architecture as discussed in Section 5.4.3, is likely to be beneficial since the origins of different tracks within a jet are to some extent correlated. Secondly, the jet classification and vertexing objectives can be considered auxiliary to the track classification task, and may bring improved track classification performance with respect to the standalone MLP.

Fig. 5.16 shows the track origin classification ROC curves for the different track origins for jets in both the $t\bar{t}$ and Z' samples. In order to improve legibility of the figure, the heavy flavour truth origins have been combined weighted by their relative abundance, as have the Primary and OtherSecondary labels. In jets in both the $t\bar{t}$ and Z' samples, the AUC of the different (grouped) origins is above 0.9, representing good classification performance. Fake tracks, followed by pileup tracks, are the easiest to classify in both samples.

		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 5.3: 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 5.4.3, outperforms the MLP model for both $t\bar{t}$ and Z' jets.

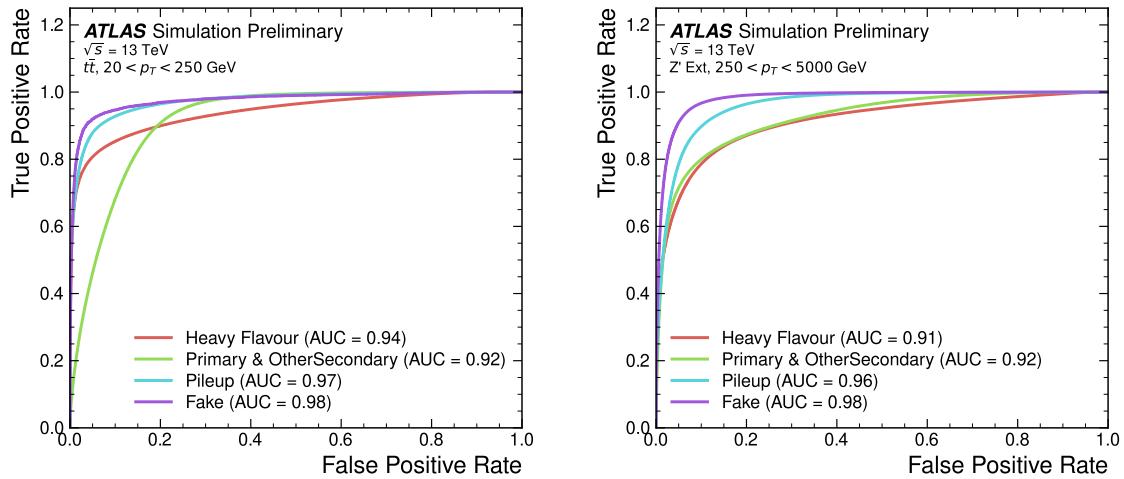


Figure 5.16: ROC curves for the different groups of truth origin labels defined in Table 4.1 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right). The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the 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.

1615 5.6 Extensions

1616 5.6.1 Looser Track Selection

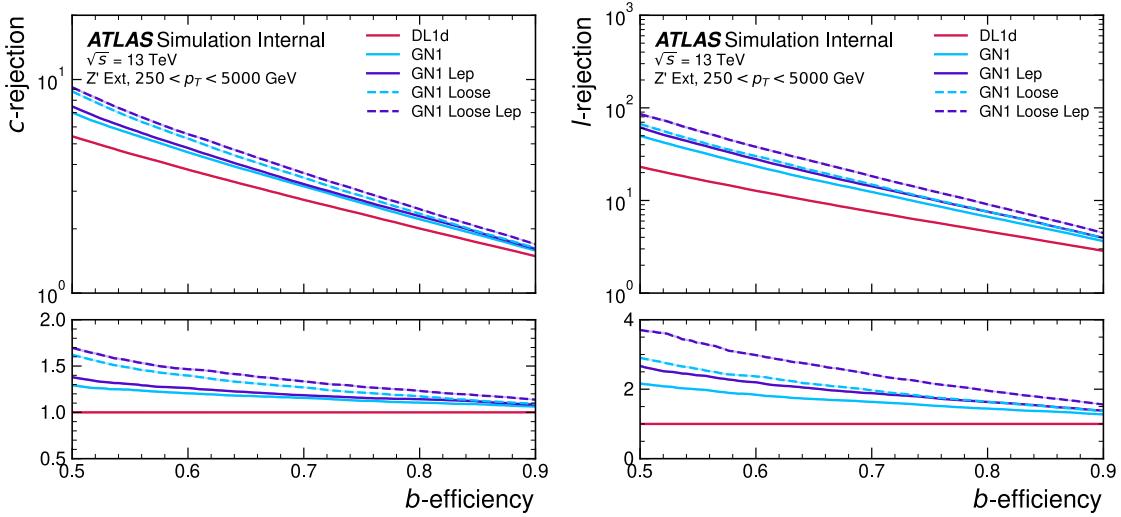


Figure 5.17: The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. A value of $f_c = 0.018$ 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.

1617 5.7 Conclusion

1618 A novel jet tagger, GN1, with a graph neural network architecture and trained
 1619 with auxiliary training targets, is presented and now fully implemented in the
 1620 ATLAS software. GN1 is shown to improve flavour tagging performance with respect
 1621 to DL1r, the current default ATLAS flavour tagging algorithm, when compared
 1622 in simulated collisions. GN1 improves c - and light-jet rejection for jets in the $t\bar{t}$
 1623 sample with $20 < p_T < 250$ GeV by factors of ~ 2.1 and ~ 1.8 respectively at a b -jet
 1624 tagging efficiency of 70% when compared with DL1r. For jets in the Z' sample
 1625 with $250 < p_T < 5000$ GeV, GN1 improves the c -rejection by a factor of ~ 2.8
 1626 and light-jet rejection by a factor of ~ 6 for a comparative b -jet efficiency of 30%.
 1627 Previous multivariate flavour tagging algorithms relied on inputs from low-level

tagging algorithms, whereas GN1 needs no such inputs, making it more flexible. It can be easily fully optimised via a retraining for specific flavour tagging use cases, as demonstrated with c -tagging and high- p_T b -tagging, without the need for time-consuming retuning of the low-level tagging algorithms. The model is also simpler to maintain and study due to the reduction of constituent components. GN1 demonstrates improved track classification performance when compared with a simple per-track MLP and an efficiency of $\sim 80\%$ for inclusive vertex finding in b -jets. The auxiliary track classification and vertex finding objectives are shown to significantly contribute to the performance in the jet classification objective, and are directly responsible for the improvement over DL1r. Further studies need to be undertaken to verify the performance of GN1 on collision data.

Future work will include adding modalities from other subdetectors and combining these in an efficient way.

₁₆₄₁

Chapter 6

₁₆₄₂

Boosted VHbb Analysis

₁₆₄₃ The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [12, 13],
₁₆₄₄ is predicted by the standard model to decay primarily to a pair of b -quarks, with a
₁₆₄₅ branching factor of 0.582 ± 0.007 [24]. Observation of this decay mode was reported by
₁₆₄₆ ATLAS [81] and CMS [26] in 2018, establishing the first direct evidence for the Yukawa
₁₆₄₇ coupling of the Higgs boson to down-type quarks (see Section 1.2.2). Whilst the
₁₆₄₈ dominant Higgs production mechanism at the LHC is gluon-gluon fusion as outlined
₁₆₄₉ in Section 1.2.3, this mechanism has an overwhelming QCD multijet background and
₁₆₅₀ so overall sensitivity to the Higgs is low. The QCD multijet background refers to
₁₆₅₁ events containing one or more strongly produced jets which are not the decay product
₁₆₅₂ of heavy resonances, for example $g \rightarrow q\bar{q}$. The gluon-gluon fusion channel contains
₁₆₅₃ to leading order only jets in the final state, and therefore it is extremely difficult to
₁₆₅₄ distinguish signal events from the overwhelming multijet background. The $H \rightarrow b\bar{b}$
₁₆₅₅ observation therefore searched for Higgs bosons produced in association with a vector
₁₆₅₆ boson V (where V can be a W or Z boson). This production mechanism results
₁₆₅₇ in leptonic final states from the decay of the vector boson, allowing for leptonic
₁₆₅₈ triggering whilst at the same time significantly reducing the multijet background.

₁₆₅₉ A closely related analysis [100] has more recently measured the associated production
₁₆₆₀ of a Higgs boson decaying to b -quarks in events where the vector and Higgs bosons are
₁₆₆₁ highly boosted. The analysis is outlined in Section 6.1. Modelling studies performed
₁₆₆₂ by the author are detailed in Section 6.2, and the results of the analysis are presented
₁₆₆₃ in Section 6.4. The author contributed to various signal and background modelling
₁₆₆₄ studies, fit studies, and to the diboson unblinding effort. This analysis has been
₁₆₆₅ published in Ref. [100]. Figures and tables from Ref. [100] are reproduced here.

cite internal
note?

1666 6.1 Analysis Overview

1667 The boosted VHbb analysis is focused on the high transverse momentum regime,
1668 which has the benefit of being more sensitive to physics beyond the Standard Model
1669 [101], but the disadvantage of being more challenging due to the increased difficulty
1670 in the accurate reconstructed of highly energy events (discussed in Chapter 3). In
1671 order to focus on the high- p_{T} regime, the reconstructed vector boson is required to
1672 have $p_{\mathrm{T}}^V > 250 \text{ GeV}$. Events are also split into two p_{T}^V bins with the first bin covering
1673 $250 \text{ GeV} < p_{\mathrm{T}}^V < 400 \text{ GeV}$ and the second covering $p_{\mathrm{T}}^V > 400 \text{ GeV}$. The full Run 2
1674 dataset at $\sqrt{s} = 13 \text{ TeV}$ is used for a total integrated luminosity of 139 fb^{-1} .

1675 The previous ATLAS analysis [81] was primarily sensitive to vector bosons with
1676 a modest p_{T}^V boost in the region of $100\text{--}300 \text{ GeV}$. The Higgs candidate was recon-
1677 structed using a pair of jets with radius parameter of $R = 0.4$, called small- R jets.
1678 However in the high- p_{T} regime, the decay products of the Higgs boson become
1679 increasingly collimated and the small- R jets may overlap. To avoid the associated
1680 problems with this and aid in the reconstruction of the Higgs boson candidate, the
1681 present analysis uses instead a large- R jet with radius parameter $R = 1.0$ [51] to
1682 reconstruct the Higgs boson candidate in all channels. The large- R jet is required
1683 to have exactly two ghost-assciated and b -tagged variable-radius track-jets, and at
1684 least one of the b -tagged track-jets is required to have a p_{T} greater than 45 GeV .
1685 The large- R jet is trimmed by removing soft and wide-angle components, which
1686 helps to remove particles from the underlying event and pileup collisions. Refer to
1687 Section 2.3.3 for more details on jet reconstruction.

1688 On top of the binning in p_{T}^V , selected events are further categorised into 0-, 1- and
1689 2-lepton channels depending on the number of selected charged leptons (electrons
1690 and muons) are present in the reconstructed final state (also referred to as 0L, 1L,
1691 and 2L respectively). The 0-lepton channel targets the $ZH \rightarrow \nu\nu b\bar{b}$ process, the
1692 1-lepton channel targets $WH \rightarrow \ell\nu b\bar{b}$, and the 2-lepton channel targets $ZH \rightarrow \ell\ell b\bar{b}$,
1693 where ℓ is an electron or muon and ν is a neutrino. Each channel has a dedicated
1694 set of selections which are listed in more detail in Section 6.1.2. The 0- and 1-lepton
1695 channels are split into high- and low-purity signal regions based on the number of
1696 additional untagged small- R jets present in the event. These channels also make use
1697 of a dedicated $t\bar{t}$ control region, described in Section 6.1.3. An complete overview of
1698 the different analysis regions is given in Table 6.1.

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

Table 6.1: 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.

1699 As mentioned, the Higgs candidate is formed from a large- R jet which contains
1700 exactly two b -tagged variable-radius track-jets jets. The track-jets jets are b -tagged
1701 using the MV2c10 b -tagging algorithm [47, 88, 102]. MV2c10 is a machine learning
1702 algorithm using a Boosted Decision Tree (BDT) which is tuned to achieve an average
1703 b -jet efficiency of 70% on simulated $t\bar{t}$ events. At this efficiency working point,
1704 rejection factors for c -jets and light-jets are approximately 9 and 304 respectively.
1705 The MV2 algorithm takes inputs from the outputs of a number of low-level algorithms
1706 (IPxD, SV1 and JetFitter). The outputs of the low-level algorithms are provided
1707 as inputs to the boosted decision tree. The efficiency of the tagging algorithm is
1708 calibrated to events in data [103–105].

1709 6.1.1 Simulated Samples

1710 How much detail?

1711 6.1.2 Selection Criteria

1712 A full list of selection cuts applied to the different analysis regions is given in Table 6.2.

1713 At least two constituent tracks with $p_T > 500$ MeV and $|\eta| < 2.5$ are required per
1714 track-jet. The track-jets themselves must satisfy $p_T > 10$ GeV and $|\eta| < 2.5$.

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 6.2: 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.

6.1.3 Signal and Control Regions

Events are further split into 2-jet or 3-jet categories depending on whether additional untagged small- R jets with $R = 0.4$ are present.

In the 0- and 1-lepton channels, the analysis is split into signal and control regions. The signal region (SR) should contain the events with Higgs candidates, while the control region (CR) is pure in $t\bar{t}$ events and hence used to constrain the normalisation of the $t\bar{t}$ background.

To leading order, there are no additional b -jets in the event other than the two coming from the reconstructed Higgs candidate, which are inside the large- R jet. For this reason, there is a signal region veto (i.e. events are not accepted into the signal region) for events with additional b -tagged jets in the event. Events with additional b -tagged jets outside the large- R jet are included in the control region. In this case, the Higgs candidate (the large- R jet containing two b -tagged track-jets) is likely to be built from the boosted hadronic decay of a top quark (i.e. $t \rightarrow Wb$). An additional b -tagged jet from the other top quark is likely to be present in the event, which is highly pure in $t\bar{t}$ events.

6.2 Modelling & Systematic Uncertainties

6.2.1 Overview

6.2.2 Sources of Systematic Uncertainties

Alternative Samples

As mentioned, alternative samples of V+jets events was generated using MADGRAPH5_AMC@NLO+PYTHIA8, and the results are compared with the nominal SHERPA 2.2.1 samples. This allows for a comparison of different parton showering and underlying event models, and derivation of the systematic uncertainties on the nominal choice of models.

Source of Uncertainty	Implementation
Renormalisation scale (μ_R)	Internal weights
Factorisation scale (μ_F)	Internal weights
PDF set	Internal weights
α_S value	Internal weights
Parton Shower (PS) models	Alternative samples
Underlying Event (UE) models	Alternative samples
Resummation scale (QSF)	Parameterisation
CKKW merging scale	Parameterisation

Table 6.3: Different sources of uncertainty (i.e. variations in the model) considered for V+jets background, and the corresponding implementation. For each uncertainty, acceptance and shape uncertainties are derived.

1740 Internal Weight Variations

1741 Nominal signal samples generated with SHERPA 2.2.1 include systematic variations
1742 of certain modelling parameters which are stored as alternative event weights. The
1743 samples contain event weight variations which correspond to variations of renormal-
1744 isation scale μ_R , and factorisation scale μ_F , of 0.5 and 2 times the nominal value.
1745 Additionally stored is event weight variations corresponding to 30 different variations
1746 on the PDF and two variations of the strong coupling constant α_S . Variations of
1747 α_S were found to have negligible impact on the results of the analysis, and are not
1748 discussed further.

1749 Parameterisation Methods

1750 While the inclusion of internal weight variation in MC event generators has decreased
1751 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently
1752 some sources of systematic uncertainty that are unable to be stored as internal weight
1753 variations due to technical limitations. Two such systematics relate to the choice of
1754 CKKW matrix element merging scale, and resummation scale (QSF). The generation
1755 of high statistics alternative samples is a time consuming process, as is typically
1756 not done for all samples for every new generator release. A method to parameterise
1757 the systematic variation using one sample, and to then apply this parameterisation
1758 to another sample, has been developed by the ATLAS SUSY group [106]. This

1759 method was used to derive CKKW and QSF uncertainties for the nominal SHERPA
1760 2.2.1 sample, using a previous (lower statistic) SHERPA 2.1 alternative sample. The
1761 resulting uncertainties were studied and found to be negligible in comparison with
1762 systemics from other sources.

1763 6.2.3 Shape Uncertainties

1764 In order to derive shape uncertainties (which as the name suggests affect shapes
1765 but not overall normalisations of distributions), the following procedure is carried
1766 out. Normalised distributions of the reconstructed Higgs candidate mass m_J are
1767 compared for the nominal sample and variations. For each variation, the ratio of the
1768 variation to nominal is calculated, and an analytic function is fit to those sources
1769 of variation which have a ratio deviating from unity. If different analysis regions or
1770 channels show the same pattern of variation, a common uncertainty is assigned. An
1771 example of a significant source of uncertainty, arising from choice of factorisation
1772 scale μ_R is shown in Fig. 6.1. An exponential function has been fitted to the ratio
1773 of the normalised distributions. Two different analysis regions (medium and high
1774 p_T^V bins) are shown. The difference of the shape of the variation means that two
1775 separate uncertainties have to be added in the fit, and applied individually in each
1776 p_T^V region.

1777 6.2.4 Acceptance Uncertainties

1778 Several different types of acceptance uncertainties have been calculated. These
1779 are implemented as nuisance parameters in the fit and for the most part account
1780 for the migration of events between different analysis regions. The list acceptance
1781 uncertainties relevant to the V+jets processes are given summarised below.

- 1782 • **Overall normalisation:** only relevant where normalisation cannot be left
1783 floating (i.e. determined in the fit).
- 1784 • **SR-to-CR relative acceptance:** the uncertainty on the normalisation of the
1785 signal region due to events migrating between the signal and control regions.

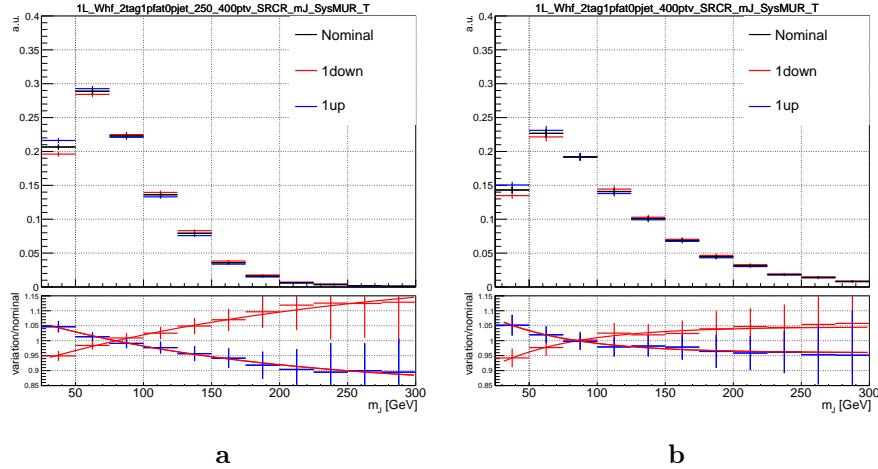


Figure 6.1: Normalised distributions of leading fat jet mass m_J for medium (6.1a) and high (6.1b) p_T^V analysis regions for W+heavy-flavour-jets in the 0 lepton channel. Merged in heavy flavours, high and low purity signal regions. The renormalisation scale μ_R has been varied by a factor of 2 (“1up”) and 0.5 (“1down”). An exponential function has been fit to the ratio.

- **HP-to-LP relative acceptance:** the uncertainty on the normalisation of the high-purity (HP) signal region due to events migrating between the high- and low-purity signal regions.
 - **Medium-to-high p_T^V relative acceptance:** describes any ‘shape’ effect in p_T^V distribution, given that the analysis only uses two p_T^V bins (medium and high).
 - **Flavour relative acceptance:** for each flavour V_{xx} , where $xx \in \{bc, bl, cc\}$ the ratio of V_{xx}/V_{bb} events is calculated. This corresponds to the uncertainty of V_{bb} events due to the miss-tagging of other flavours V_{xx} .
- The uncertainties on different systematics are summed in quadrature to give a total uncertainty on each region. A summary of the different acceptance uncertainties that were derived in this way and subsequently applied in the fit are given in Table 6.4. An effort has been made, wherever possible, to harmonise similar uncertainties across different analysis regions and channels.

1799 6.2.5 Vector Boson + Jets Modelling

1800 The background processes involving W or Z boson decays into leptons (including
1801 those in which the W boson arises from a top-quark decay) are collectively referred
1802 to as electroweak (EW), or V+jets, backgrounds. W + jets events are most relevant
1803 to the 1-lepton channel via the leptonic decay of $W \rightarrow \ell\nu$. In the event of $W \rightarrow \tau\nu$,
1804 and subsequent decay of the τ , or the lack of the successful reconstruction of the
1805 e or μ , W + jets can also contribute to the 0-lepton channel. Meanwhile, Z+jets
1806 contributes primarily to the 0- and 2-lepton channels via the processes $Z \rightarrow \nu\nu$ and
1807 $Z \rightarrow \ell\ell$ respectively.

1808 Modelling is used to predict the outcomes of the analysis and to assess the impact of
1809 sources of different systematic uncertainty. Signal and background modelling has
1810 has primarily consisted of using Monte Carlo (MC) generators to produce simulated
1811 events. The uncertainties on the simulated output must be well understood to
1812 perform a successful analysis. To achieve this, a set of “nominal” samples are first
1813 defined as a reference to which different variations can be compared. The nominal
1814 samples are chosen as the best possible representation of the underlying physical
1815 process. “Alternative” samples are used to understand the systematic uncertainties
1816 on the nominal samples. To generate an alternative sample, some aspect of the model
1817 is varied, and the simulation is re-run. A comparison back to the nominal sample
1818 gives a handle on the systematic uncertainty associated with the model parameter
1819 which was changed. Detailed information can be found in [107]. In order to access
1820 uncertainties associated with the use of MC generators, variations of the data are
1821 produced using alternative generators or variation of nominal generator parameters.
1822 The variation of nominal generator parameters can in certain cases be implemented
1823 using internal weight variations stored alongside the nominal events, and in other
1824 cases a new independent sample must be generated. The nominal generator used
1825 for V+jets events is SHERPA 2.2.1, while MADGRAPH5_AMC@NLO+PYTHIA8
1826 (which uses different parton showering models) is used as an alternative generator.
1827 As production of large MC samples is computationally expensive, a feature of state
1828 of the art simulation packages is to store some sources of variation as internal event
1829 weights, which can be generated alongside the nominal samples, saving computation
1830 time. Several sources of uncertainty, summarised in Table 6.3, have been assessed.

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

1831 **6.2.6 Diboson Modelling**

1832 **6.3 Fit Model**

1833 A global profile likelihood fit is used to extract the signal strength μ and its significance
1834 from the data. This statistical setup treats each bin as a Poisson counting experiment.
1835 The combined likelihood over N bins, without considering sources of systematic
1836 uncertainty, is given by

$$\mathcal{L}(\mu) = \prod_{i=1}^N \frac{(\mu s_i + b_i)^{n_i}}{n_i!} \exp [-(\mu s_i + b_i)], \quad (6.1)$$

1837 where s_i (b_i) is the expected number of signal (background) events in bin i , and
1838 n_i is the number of events observed in data in bin i . The presence of systematic
1839 uncertainties which can affect the expected numbers of signal and background events
1840 necessitates the addition of nuisance parameters (NPs), θ , to the likelihood. Each
1841 source of systematic uncertainty for V+jets samples discussed in the previous section
1842 was implemented as a NP θ_j in the fit. The presence of NPs modifies the likelihood
1843 as

$$\mathcal{L}(\mu) \rightarrow \mathcal{L}(\mu, \theta) = \mathcal{L}(\mu) \times \mathcal{L}(\theta), \quad s_i \rightarrow s_i(\theta), \quad b_i \rightarrow b_i(\theta), \quad (6.2)$$

1844 where

$$\mathcal{L}(\theta) = \prod_{\theta_j \in \theta} \frac{\exp [-\theta_j^2/2]}{\sqrt{2\pi}}. \quad (6.3)$$

1845 **6.4 Results**

1846 **6.4.1 Pre-fit Yields**

1847 **6.4.2 Post-fit Results**

1848 Post-fit m_J distributions in the high-purity medium p_T^V regions for the 0- and 2-
1849 lepton channels are shown in Fig. 6.2. The plots show large falling backgrounds,

1850 predominantly made up of W +jets and Z +jets events, and a signal distribution
1851 corresponding to the Standard Model Higgs boson peaking around $m_H = 125$ GeV.

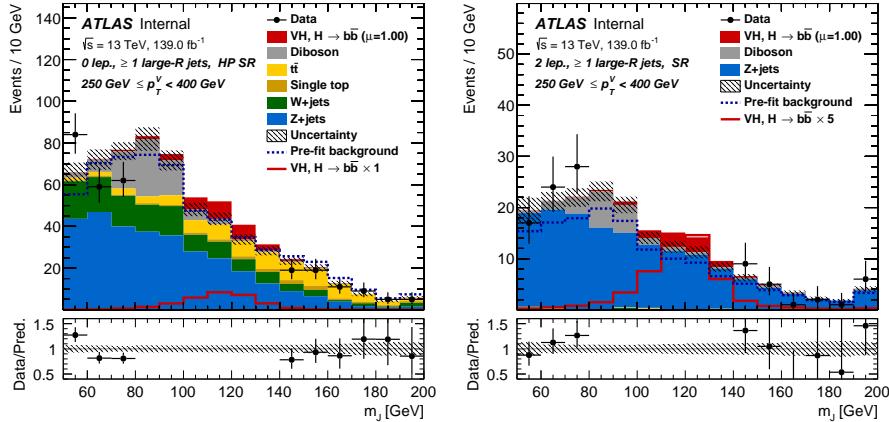


Figure 6.2: Post-fit distributions for the 0-lepton (left) and 2-lepton (right) channels in the high purity medium p_T^V region, obtained in the combined conditional $\mu = 1$ fit to data. The last bin of each plot is an overflow bin.

1852 6.5 Improved b -tagging using GNNs

1853 6.6 Conclusion

1854 Work has been carried out as part of the boosted VHbb analysis group to understand,
1855 and implement in the global profile likelihood fit, systematic uncertainties on V+jets
1856 samples. This background modelling work is an essential part of the success of
1857 the analysis. So far the fit has proved stable with the inclusion of the V+jets
1858 uncertainties, and detailed studies are now underway to determine the causes behind
1859 any observed pulls of the added NPs. Additional work is ongoing to help with the
1860 derivation of uncertainties on diboson samples, another important background. The
1861 analysis is already advanced, and is now progressing into its final stages. Publication
1862 is expected in the new year.

1863 This analysis would benefit greatly from the improved high p_T b -tagging enabled by
1864 GN1.

Source of uncertainty	Avg. impact	
Total	0.372	
Statistical	0.283	
Systematic	0.240	
Experimental uncertainties		
Small- R jets	0.038	
Large- R jets	0.133	
E_T^{miss}	0.007	
Leptons	0.010	
b -tagging	b -jets	0.016
	c -jets	0.011
	light-flavour jets	0.008
	extrapolation	0.004
Pile-up	0.001	
Luminosity	0.013	
Theoretical and modelling uncertainties		
Signal	0.038	
Backgrounds	0.100	
$\hookrightarrow Z + \text{jets}$	0.048	
$\hookrightarrow W + \text{jets}$	0.058	
$\hookrightarrow t\bar{t}$	0.035	
\hookrightarrow Single top quark	0.027	
\hookrightarrow Diboson	0.032	
\hookrightarrow Multijet	0.009	
MC statistical	0.092	

Table 6.5: Breakdown of the absolute contributions to the uncertainty in μ_{VH}^{bb} inclusive in p_T^V . The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations. The reported values represent the average between the positive and negative uncertainties on μ_{VH}^{bb} .

₁₈₆₅ Chapter 7

₁₈₆₆ Conclusion

₁₈₆₇ Summary, tie it up, future work.

₁₈₆₈ **Bibliography**

- ₁₈₆₉ [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- ₁₈₇₀
- ₁₈₇₁ [2] L. Morel, Z. Yao, P. Cladé and S. Guellati-Khélifa, *Determination of the fine-structure constant with an accuracy of 81 parts per trillion*, *Nature* **588** (2020) pp. 61–65.
- ₁₈₇₂
- ₁₈₇₃
- ₁₈₇₄ [3] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al., *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* **606** (2022) pp. 479–483.
- ₁₈₇₅
- ₁₈₇₆
- ₁₈₇₇ [4] CDF collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, *Phys. Rev. Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- ₁₈₇₈
- ₁₈₇₉ [5] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) pp. 2632–2637 [[hep-ex/9503003](#)].
- ₁₈₈₀
- ₁₈₈₁ [6] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) pp. 252–255.
- ₁₈₈₂
- ₁₈₈₃ [7] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, *Phys. Lett. B* **122** (1983) pp. 103–116.
- ₁₈₈₄
- ₁₈₈₅
- ₁₈₈₆ [8] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) pp. 218–224 [[hep-ex/0012035](#)].
- ₁₈₈₇
- ₁₈₈₈ [9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- ₁₈₈₉
- ₁₈₉₀ [10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) pp. 508–509.
- ₁₈₉₁
- ₁₈₉₂ [11] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*

- 1893 *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 1894 [12] ATLAS Collaboration, *Observation of a new particle in the search for the Standard
1895 Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)
1896 p. 1 [[1207.7214](#)].
- 1897 [13] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS
1898 experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 1899 [14] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)
1900 p. 083C01.
- 1901 [15] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge
1902 invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 1903 [16] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)
1904 pp. 579–588.
- 1905 [17] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 1906 [18] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel
1907 symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**
1908 (1968) pp. 367–377.
- 1909 [19] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,
1910 *Phys. Rev.* **104** (1956) pp. 254–258.
- 1911 [20] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental
1912 test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 1913 [21] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of
1914 conservation of parity and charge conjugation in meson decays: the magnetic moment
1915 of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 1916 [22] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,
1917 *Phys. Rev.* **117** (1960) pp. 648–663.
- 1918 [23] J. Goldstone, *Field theories with «superconductor» solutions*, *Il Nuovo Cimento
1919 (1955-1965)* **19** (1961) pp. 154–164.
- 1920 [24] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs
1921 Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](#).
- 1922 [25] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*

- 1923 *ATLAS detector*, ATLAS-CONF-2018-036 (2018).
- 1924 [26] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys. Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](#)].
- 1925 [27] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) pp. S08002–S08002.
- 1926 [28] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008)
- 1927 p. S08004.
- 1928 [29] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of Instrumentation* **3** (2008) pp. S08005–S08005.
- 1929 [30] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) p. S08003.
- 1930 [31] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 1931 [32] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 1932 [33] 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](#).
- 1933 [34] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, *JINST* **11** (2016) p. P04008 [[1512.01094](#)].
- 1934 [35] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 1935 [36] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 1936 [37] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 1937 [38] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) p. T05008 [[1803.00844](#)].

- 1953 [39] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) p. 317 [[1611.09661](#)].
- 1954 [40] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, (2021).
- 1956 [41] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 1958 [42] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense Environments*, ATL-PHYS-PUB-2015-006 (2015).
- 1960 [43] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) p. 673 [[1704.07983](#)].
- 1962 [44] ATLAS collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, arXiv e-prints (2014) p. arXiv:1406.7690 [[1406.7690](#)].
- 1964 [45] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332 [[1611.10235](#)].
- 1967 [46] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, ATL-PHYS-PUB-2019-015 (2019).
- 1969 [47] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **79** (2019) p. 970 [[1907.05120](#)].
- 1972 [48] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) p. 063 [[0802.1189](#)].
- 1974 [49] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 1976 [50] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].
- 1979 [51] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008) p. 242001 [[0802.2470](#)].
- 1982 [52] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012).

- 1984 [53] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*
1985 *using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024 (2016).
- 1986 [54] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000)
1987 pp. 577–606 [[hep-ph/9912292](#)].
- 1988 [55] A. Chen, M. Goldberg, N. Horwitz, A. Jawahery, P. Lipari, G. C. Moneti et al.,
1989 *Limit on the $b \rightarrow u$ coupling from semileptonic b decay*, *Phys. Rev. Lett.* **52** (1984)
1990 pp. 1084–1088.
- 1991 [56] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).
1992 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2022-03/>.
- 1993 [57] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS*
1994 *Experiment*, .
- 1995 [58] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*
1996 *Process Interface for Event Generators*, *ArXiv High Energy Physics - Phenomenology*
1997 *e-prints* (2001) .
- 1998 [59] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*
1999 *standard format for Les Houches Event Files*, *Computer Physics Communications*
2000 **176** (2007) pp. 300–304.
- 2001 [60] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*
2002 *universal approximators*, *Neural Networks* **2** (1989) pp. 359–366.
- 2003 [61] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*
2004 *back-propagating errors*, *nature* **323** (1986) pp. 533–536.
- 2005 [62] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*
2006 *activity*, *The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.
- 2007 [63] J. Hopfield, *Neural networks and physical systems with emergent collective*
2008 *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*
2009 *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 2010 [64] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*
2011 *networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 2012 [65] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*
2013 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2014 [66] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,

- 2015 *Journal of High Energy Physics* **2004** (2004) p. 040–040.
- 2016 [67] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* **2007** (2007) p. 126–126.
- 2019 [68] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton shower simulations: the powheg method*, *Journal of High Energy Physics* **2007** (2007) p. 070–070.
- 2022 [69] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy Physics* **2010** (2010) .
- 2025 [70] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015) p. 040 [[1410.8849](#)].
- 2027 [71] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020 (2016).
- 2029 [72] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159 [[1410.3012](#)].
- 2032 [73] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021 (2014).
- 2034 [74] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) p. 244 [[1207.1303](#)].
- 2036 [75] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) p. 152.
- 2038 [76] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) p. 823 [[1005.4568](#)].
- 2040 [77] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) p. 250.
- 2042 [78] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014 (2020).
- 2044 [79] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-027 (2022).

- 2046 [80] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 2047 [81] ATLAS Collaboration, *Observation of $H \rightarrow b\bar{b}$ decays and VH production with the*
2048 *ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].
- 2049 [82] ATLAS Collaboration, *Observation of Higgs boson production in association with a*
2050 *top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173
2051 [[1806.00425](#)].
- 2052 [83] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*
2053 *using 139 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *JHEP* **03**
2054 (2020) p. 145 [[1910.08447](#)].
- 2055 [84] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**
2056 (2018) p. 030001.
- 2057 [85] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,
2058 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,
2059 *arXiv preprint arXiv:1806.01261* (2018) .
- 2060 [86] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,
2061 *Secondary vertex finding in jets with neural networks*, *The European Physical Journal*
2062 *C* **81** (2021) .
- 2063 [87] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,
2064 *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. arXiv:2002.08772
2065 [[2002.08772](#)].
- 2066 [88] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*
2067 *b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017).
- 2068 [89] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*
2069 *ATLAS detector*, ATL-PHYS-PUB-2017-011 (2017).
- 2070 [90] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
2071 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003 (2017).
- 2072 [91] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS*
2073 *detector*, ATLAS-CONF-2014-018 (2014).
- 2074 [92] ATLAS Collaboration, *Muon reconstruction performance in early $\sqrt{s} = 13\text{ TeV}$ data*,
2075 ATL-PHYS-PUB-2015-037 (2015).
- 2076 [93] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*

- 2077 *experiment using the 2015 and 2016 LHC proton–proton collision data at*
2078 $\sqrt{s} = 13\text{ TeV}$, *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 2079 [94] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*
2080 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, *arXiv e-prints* (2020)
2081 p. arXiv:2007.08294 [[2007.08294](#)].
- 2082 [95] J. Shlomi, P. Battaglia and J.-R. Vlimant, *Graph neural networks in particle physics,*
2083 *Machine Learning: Science and Technology* **2** (2021) p. 021001.
- 2084 [96] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,
2085 *arXiv e-prints* (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 2086 [97] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,
2087 *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 2088 [98] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*
2089 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2090 [99] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).
2091 <https://github.com/onnx/onnx>.
- 2092 [100] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson*
2093 *decaying into b-quarks with a vector boson at high transverse momentum in pp*
2094 *collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *Phys. Lett. B* **816** (2021)
2095 p. 136204 [[2008.02508](#)].
- 2096 [101] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated*
2097 *Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039
2098 [[1512.02572](#)].
- 2099 [102] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in*
2100 *Run-2*, ATL-PHYS-PUB-2015-022 (2015).
- 2101 [103] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS*
2102 *detector using $t\bar{t}$ events at $\sqrt{s} = 13\text{ TeV}$* , *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 2103 [104] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using*
2104 *ATLAS proton–proton collision data at $\sqrt{s} = 13\text{ TeV}$* , ATLAS-CONF-2018-006
2105 (2018).
- 2106 [105] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using*
2107 *a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001 (2018).

- 2108 [106] J. K. Anders and M. D'Onofrio, *V+Jets theoretical uncertainties estimation via a*
2109 *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 2110 [107] A. S. Bell and F. Lo Sterzo, *Signal and Background Modelling Studies for the*
2111 *Standard Model VH, $H \rightarrow b\bar{b}$ Analysis*, Tech. Rep. ATL-COM-PHYS-2018-505,
2112 Geneva (2018).

2113 List of figures

<small>2114</small> 1.1	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.	15
<small>2119</small> 1.2	Higgs 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 the Higgs mass $m_H = 125$ GeV, Higgs production is dominated by gluon-gluon fussion, vector boson fusion, and associated production with vector bosons.	21
<small>2123</small> 1.3	Higgs branching ratios as a function of Higgs mass m_H at $\sqrt{s} = 13$ TeV [24]. Uncertainties are shown in the shaded bands. At the Higgs mass $m_H = 125$ GeV, the Higgs predominantly decays to a pair of b -quarks, around 60% of the time. The subdominant decay mode is off shell to a pair of W bosons.	21
<small>2128</small> 2.1	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. [33].	25
<small>2132</small> 2.2	The track parameterisation used at the ATLAS detector. Five coordinates $(d_0, z_0, \phi, \theta, q/p)$ are specified, defined at the point closest approach of the track to the nominal interaction point at the origin of the coordinate system.	26
<small>2135</small> 2.3	Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [32]. A total of 139 fb^{-1} of collision data is labelled as good for physics.	28

2138	2.4 Average pile-up profiles measured by ATLAS during Run 2 [32]. During		
2139	Run 3, even higher levels of pile-up are expected.	29	
2140	2.5 A 3D model of the entire ATLAS detector. The detector is 46 m long and		
2141	25 m in diameter. Cutouts to the centre of the detector reveal the different		
2142	subdetectors which are arranged in layers around the nominal interaction		
2143	point.	30	
2144	2.6 A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors,		
2145	showing the barrel layers and end-cap disks.	31	
2146	2.7 A cross-sectional view of the ATLAS ID, with the radii of the different		
2147	barrel layers shown.	31	
2148	2.8 A schematic cross-sectional view of the ATLAS IBL.	32	
2149	2.9 The ATLAS calorimeters. The ECal (orange) and HCal (grey, brown). . .	34	
2150	2.10 The ATLAS muon spectrometer.	36	
2151	2.11 Particles (left) which have enough separation will leave charge depositions		
2152	which are resolved into separate clusters. Sufficiently close particles (right)		
2153	can lead to merged clusters. Their combined energy deposits are recon-		
2154	structed as a single merged cluster.	39	
2155	2.12 A sketch of electron reconstruction using the ATLAS detector [53]. Electron		
2156	reconstruction makes use of the entire ID and the calorimeters. In particular,		
2157	discriminating information comes from the TRT and ECal.	42	
2158	3.1 Diagram of a typical b -jet (blue) which has been produced in an event		
2159	alongside two light jets (grey). The b -hadron has travelled a significant		
2160	distance (pink dashed line) from the primary interaction point (pink dot)		
2161	before its decay. The large transverse impact parameter d_0 is a characteristic		
2162	property of the trajectories of b -hadron decay products.	45	
2163	3.2 The truth b -hadron decay radius L_{xy} as a function of truth transverse		
2164	momentum p_T for reconstructed b -jets in Z' events. Error bars show the		
2165	standard deviation. The pixel layers are shown in dashed horizontal lines.	46	

2166	3.3 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, instead reconstructing merged clusters.	47
2172	3.4 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. For each of these, standard tracks and pseudo-tracks are plotted. Hit multiplicities on the pseudo-tracks at high p_T due to the increased flight of the b -hadron. The baseline tracks have more hits than the pseudo-tracks, indicating that they are being incorrectly assigned additional hits.	48
2179	3.5 b -hadron decay track reconstruction efficiency as a function of truth b -hadron p_T [56]. 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.	49
2186	3.6 The rate of shared hits 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. Pseudo-tracks represent ideal performance given the ATLAS detector, and given also the efficiency of the NN cluster splitting algorithms.	52
2190	3.7 Profiles, as a function of parent b -hadron p_T , of good (left) and wrong (right) hit assignment rates on the IBL for tracks using baseline tracking (black), the modified version of the outlier removal procedure (red).	54
2193	3.8 (left) B hadron decay track d_0 pulls for baseline and modified GX2F tracks. (right) The magnitude of the decay track d_0 pull as a function of B hadron transverse momentum.	55
2196	4.1 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 is passed to an activation function. Each neuron can be thought of as a logistic regression model.	58

2200	4.2 Several common choices for the activation function g of an artificial neuron.	59
2201	4.3 Rate of fake tracks as a function of jet transverse momentum (left) and	
2202	$\Delta R(\text{track}, \text{jet})$ (right). The rate of fake tracks increases significantly as a	
2203	function of p_T , and also increases as the distance to the jet axis decreases,	
2204	and the number of tracks in the jet increases (not shown).	62
2205	4.4 The light-jet efficiency of the low level tagger SV1 as a function of b -jet	
2206	efficiency for the nominal tracking setup (black) and for the case where	
2207	fake tracks which are not from the decay of a b -hadron are removed. The	
2208	light-jet efficiency is decreased, demonstrating that the presence of fake	
2209	tracks is detrimental to algorithm performance.	63
2210	4.5 (left) Normalised histogram of the model output separated for good and	
2211	fake tracks, and further separated by those tracks which are from the decay	
2212	of a b -hadron. (right) The ROC curve for all tracks (solid line) and tracks	
2213	from the decay of a b -hadron (dashed line).	66
2214	4.6 The effect of applying the fake track identification algorithm alongside	
2215	the b -hadron decay track identification on the jet tagging performance	
2216	of SV1 for jets with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and for jets with	
2217	$400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black)	
2218	is compared to two working points of fake track removal (orange and red).	
2219	Removal of fake tracks based on truth information is shown by the green	
2220	curves.	68
2221	5.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right).	
2222	The existing approach utilises low-level algorithms (shown in blue), the	
2223	outputs of which are fed into a high-level algorithm (DL1r). Instead	
2224	of being used to guide the design of the manually optimised algorithms,	
2225	additional truth information from the simulation is now being used as	
2226	auxiliary training targets for GN1. The solid lines represent reconstructed	
2227	information, whereas the dashed lines represent truth information.	71
2228	5.2 The inputs to GN1 are the two jet features ($n_{\text{jf}} = 2$), and an array of	
2229	n_{tracks} , where each track is described by 21 track features ($n_{\text{tf}} = 21$). The	
2230	jet features are copied for each of the tracks, and the combined jet-track	
2231	vectors of length 23 form the inputs of GN1.	77

2232	5.3	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.	78
2238	5.4	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.	82
2244	5.5	The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. 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.	84
2253	5.6	The c -jet (left) and light-jet (right) rejections as a function of the b -jet tagging efficiency for jets in the Z' sample with $250 < p_T < 5000$ GeV. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b for DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. 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.	85
2262	5.7	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 GN1 Lep. Binomial error bands are denoted by the shaded regions.	87

2267	5.8 The b -jet (left) and light-jet (right) rejections as a function of the c -jet 2268 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV. The ratio to the 2269 performance of the DL1r algorithm is shown in the bottom panels. Binomial 2270 error bands are denoted by the shaded regions. At c -jet tagging efficiencies 2271 than $\sim 25\%$, the light-jet rejection becomes so large that the effect of the 2272 low number of jets is visible. The lower x -axis range is chosen to display 2273 the c -jet tagging efficiencies usually probed in these regions of phase space.	89
2274	5.9 The b -jet (left) and light-jet (right) rejections as a function of the c -jet 2275 tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV. The ratio to the 2276 performance of the DL1r algorithm is shown in the bottom panels. Binomial 2277 error bands are denoted by the shaded regions. The lower x -axis range 2278 is chosen to display the c -jet tagging efficiencies usually probed in these 2279 regions of phase space.	89
2280	5.10 A schematic showing the truth track origin and vertex information (left) 2281 compared with the predictions from GN1 (right). Vertices reconstructed by 2282 SV1 and JetFitter are also marked.	90
2283	5.11 A schematic showing the truth track origin and vertex information (left) 2284 compared with the predictions from GN1 (right). Vertices reconstructed by 2285 SV1 and JetFitter are also marked.	90
2286	5.12 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 2287 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal 2288 GN1, in addition to configurations where no (GN1 No Aux), only the track 2289 classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives 2290 are deployed. The ratio to the performance of the DL1r algorithm is shown 2291 in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b 2292 for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted 2293 by the shaded regions. At b -jet tagging efficiencies less than $\sim 65\%$, the 2294 light-jet rejection become so large that the effect of the low number of jets 2295 are visible. The lower x -axis range is chosen to display the b -jet tagging 2296 efficiencies usually probed in these regions.	91

2297	5.13 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 2298 tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the nominal 2299 GN1, in addition to configurations where no (GN1 No Aux), only the track 2300 classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives 2301 are deployed. The ratio to the performance of the DL1r algorithm is shown 2302 in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of D_b 2303 for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted 2304 by the shaded regions. At b -jet tagging efficiencies less than $\sim 25\%$, the 2305 light-jet rejection become so large that the effect of the low number of jets 2306 are visible. The lower x -axis range is chosen to display the b -jet tagging 2307 efficiencies usually probed in these regions.	92
2308	5.14 Vertex finding efficiency as a function of jet p_T for b -jets in the $t\bar{t}$ sample 2309 using the exclusive (left) and inclusive (right) vertex finding approaches. 2310 Efficient vertex finding requires the recall of at least 65% of the tracks in 2311 the truth vertex, and allows no more than 50% of the tracks to be included 2312 incorrectly. Binomial error bands are denoted by the shaded regions.	95
2313	5.15 Inclusive vertex finding efficiency for multitrack truth vertices in b -jets in 2314 the $t\bar{t}$ sample (left) and jets in the Z' sample (right) as a function of jet p_T . 2315 Efficient vertex finding requires the recall of at least 65% of the tracks in 2316 the truth vertex, and allows no more than 50% of the tracks to be included 2317 incorrectly.	96
2318	5.16 ROC curves for the different groups of truth origin labels defined in Table 4.1 2319 for jets in the $t\bar{t}$ sample (left) and jets in the Z' sample (right). The 2320 FromB, FromBC and FromC labels have been combined, weighted by their 2321 relative abundance, into the Heavy Flavour category, and the Primary and 2322 OtherSecondary labels have similarly been combined into a single category. 2323 The mean weighted area under the ROC curves (AUC) is similar for both 2324 samples.	98
2325	5.17 The c -jet (left) and light-jet (right) rejections as a function of the b -jet 2326 tagging efficiency for Z' jets with $250 < p_T < 5000$ GeV, for the looser 2327 track selection trainings of GN1. The ratio to the performance of the DL1d 2328 algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in 2329 the calculation of D_b for DL1d and $f_c = 0.05$ is used for GN1. Binomial 2330 error bands are denoted by the shaded regions.	99

2331	6.1 Normalised distributions of leading fat jet mass m_J for medium (6.1a) and	
2332	high (6.1b) p_T^V analysis regions for W+heavy-flavour-jets in the 0 lepton	
2333	channel. Merged in heavy flavours, high and low purity signal regions. The	
2334	renormalisation scale μ_R has been varied by a factor of 2 (“1up”) and 0.5	
2335	(“1down”). An exponential function has been fit to the ratio.	108
2336	6.2 Post-fit distributions for the 0-lepton (left) and 2-lepton (right) channels in	
2337	the high purity medium p_T^V region, obtained in the combined conditional	
2338	$\mu = 1$ fit to data. The last bin of each plot is an overflow bin.	112

2339 List of tables

<small>2340</small>	1.1	The half-integer spin 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>2341</small>	10				
<small>2342</small>								
<small>2343</small>	1.2	The integer spin 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 particle is the only scalar boson in the SM.	<small>2344</small>	<small>2345</small>	11			
<small>2346</small>								
<small>2347</small>	2.1	Overview of the different LHC runs [31, 32]. The average number of interactions per bunch-crossing is denoted as $\langle \mu \rangle$ (see Section 2.1.3). Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.	<small>2348</small>	<small>2349</small>	<small>2350</small>	24		
<small>2351</small>								
<small>2352</small>	4.1	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 [43]. 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>2353</small>	<small>2354</small>	<small>2355</small>	<small>2356</small>	<small>2357</small>	61
<small>2358</small>								
<small>2359</small>	4.2	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 [43], while split hits are hits used on multiple tracks which have been identified as merged.	<small>2360</small>	<small>2361</small>	<small>2362</small>	<small>2363</small>	65	
<small>2364</small>								

2364	4.3 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. Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [43]. Shared hits on pixel layers are given a weight of 1, while shared hits in the SCT are given a weight of 0.5. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.	66
2371	4.4 Good and fake track selection efficiencies for the combined $t\bar{t}$ and Z' samples. Two working points are defined, cutting on the NN output at 0.06 and 0.12. The continuous output of the model allows for the tuning of good and fake track identification efficiencies.	67
2375	5.1 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 [43], 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 GN1 Lep model.	75
2382	5.2 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.	79
2385	5.3 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 5.4.3, outperforms the MLP model for both $t\bar{t}$ and Z' jets.	98
2392	6.1 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.	103

2398	6.2 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.	104
2402	6.3 Different sources of uncertainty (i.e. variations in the model) considered for $V+jets$ background, and the corresponding implementation. For each uncertainty, acceptance and shape uncertainties are derived.	106
2405	6.4 $V+jets$ 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.	110
2413	6.5 Breakdown of the absolute contributions to the uncertainty in μ_{VH}^{bb} inclusive in p_T^V . The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations. The reported values represent the average between the positive and negative uncertainties on μ_{VH}^{bb}	113