

# 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 2, 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

Here some useful packages are demonstrated. In particular, the `hepunit` package which adds additional units to `SIUnit`. A variety of jet measurements are made using data collected during the first year of 7 TeV proton-proton collisions from the general-purpose ATLAS experiment at the LHC. no more than 300 words

# **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<sup>A</sup>T<sub>E</sub>X 2 <sub>$\varepsilon$</sub>  using the “heptesis” class [1].

Brian Moser provided insight on the fit.

# Contents

<b>1</b>	<b>Theoretical Framework</b>	<b>9</b>
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
<b>2</b>	<b>The Large Hadron Collider and the ATLAS Detector</b>	<b>23</b>
2.1	Coordinate System & Collider Definitions . . . . .	24
2.1.1	ATLAS Coordinate System . . . . .	24
2.1.2	Collider Definitions . . . . .	25
2.2	The ATLAS Detector . . . . .	26
2.2.1	The Inner Detector . . . . .	28
2.2.2	Calorimeters . . . . .	31
2.2.3	Muon Spectrometer . . . . .	32
2.2.4	The Trigger . . . . .	33
2.3	Reconstructed Physics Objects . . . . .	34
2.3.1	Tracks . . . . .	34
2.3.2	Vertices . . . . .	36
2.3.3	Jets . . . . .	38
2.3.4	Leptons . . . . .	39
<b>3</b>	<b>Tracking and <math>b</math>-tagging</b>	<b>41</b>
3.1	$b$ -hadron Reconstruction . . . . .	41
3.1.1	Decay Topology . . . . .	41

3.1.2	Challenges Facing $b$ -hadron Reconstruction . . . . .	44
3.2	Investigating Improvements for High $p_T$ B Tracking . . . . .	46
3.2.1	Pseudotrack and Ideal Tracks . . . . .	46
3.2.2	Global $\chi^2$ Fitter Outlier Removal . . . . .	47
3.3	Tracking software validation . . . . .	51
<b>4</b>	<b>Track Classification MVA</b>	<b>52</b>
4.1	Machine Learning Background . . . . .	52
4.1.1	Neural Networks . . . . .	54
4.1.2	Training with Gradient Descent . . . . .	55
4.2	Track Truth Origin Labelling . . . . .	56
4.3	Fake Track Identification Tool . . . . .	57
4.3.1	Model Inputs . . . . .	59
4.3.2	Results . . . . .	59
4.4	Conclusion . . . . .	59
<b>5</b>	<b>Graph Neural Network Flavour Tagger</b>	<b>63</b>
5.1	Motivation . . . . .	64
5.2	Graph Neural Network Theory . . . . .	67
5.3	Experimental Setup . . . . .	67
5.3.1	Datasets . . . . .	67
5.4	Model Architecture . . . . .	68
5.4.1	Model Inputs . . . . .	68
5.4.2	Auxiliary Training Objectives . . . . .	70
5.4.3	Architecture . . . . .	71
5.4.4	Training . . . . .	74
5.5	Results . . . . .	76
5.5.1	$b$ -tagging Performance . . . . .	77
5.5.2	$c$ -tagging Performance . . . . .	82
5.5.3	Ablations . . . . .	82
5.5.4	Inclusion of Low-Level Vertexing Algorithms . . . . .	84
5.5.5	Vertexing Performance . . . . .	87
5.5.6	Track Classification Performance . . . . .	90
5.6	Conclusion . . . . .	92
5.7	Extensions . . . . .	93
5.7.1	Looser Track Selection . . . . .	93

<b>6 Boosted VHbb Analysis</b>	<b>94</b>
6.1 Overview . . . . .	94
6.2 Introduction . . . . .	95
6.3 Modelling Work . . . . .	97
6.3.1 Background . . . . .	97
6.3.2 Vector Boson + Jets Modelling . . . . .	100
6.3.3 Diboson Modelling . . . . .	102
6.4 Fit Studies . . . . .	102
6.4.1 Fit Model . . . . .	102
6.5 Conclusion . . . . .	103
<b>7 Conclusion</b>	<b>104</b>
<b>A Combining Multiple Triggers</b>	<b>105</b>
<b>Bibliography</b>	<b>106</b>
<b>List of figures</b>	<b>115</b>
<b>List of tables</b>	<b>122</b>

# <sup>2</sup> Chapter 1

## <sup>3</sup> Theoretical Framework

<sup>4</sup> The Standard Model (SM) of particle physics is the theory describing all known  
<sup>5</sup> elementary particles and their interactions via three of the four fundamental forces.

<sup>6</sup> Developed by merging the successful theories of quantum mechanics and relativity  
<sup>7</sup> in the second half of the 20th century, the SM's position today at the centre of our  
<sup>8</sup> understanding of the nature of the universe is firmly established by an unparalleled  
<sup>9</sup> level of agreement between the predictions from the model and experimental results  
<sup>10</sup> [2, 3].

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

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

## <sup>23</sup> 1.1 The Standard Model

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

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

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

<sup>36</sup> The SM provides a mathematical description of how the four fundamental forces  
<sup>37</sup> 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.

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

**Table 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	$\gamma$	$< 1 \times 10^{-27}$	$< 1 \times 10^{-46}$	1
Weak boson	$W^\pm$	$80.377 \pm 0.012$	$\pm 1$	1
Weak boson	$Z$	$91.1876 \pm 0.0021$	0	1
Gluon	$g$	0	0	1
Higgs	$H$	$125.25 \pm 0.17$	0	0

**Table 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  $\psi^\dagger$  denotes the Hermitian conjugate of  $\psi$ .  
 43 The field  $\psi$  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  $\gamma^\mu$  (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  $\alpha$  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  $A_\mu$  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  $\psi$  and  
 57  $A_\mu$  to Eq. (1.2). For completeness, the kinetic term for the the new field  $A_\mu$  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  $\psi$ . 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  $A_\mu$  must be  
 63 massless. Requiring invariance under local  $U(1)$  gauge transformations necessitated  
 64 the addition of a new field  $A_\mu$ , 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  $\lambda_a$  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

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

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

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

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

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

The full QCD Lagrangian is then given by

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

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

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

### 1.1.3 The Electroweak Sector

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

- 101 The generators of  $SU(2)_L$  are  $T_a = \sigma_a/2$ , where  $\sigma_a$  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?

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

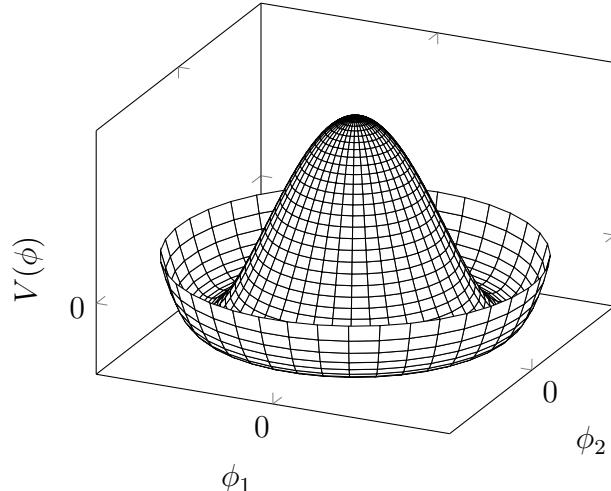
## 120 1.2 The Higgs Mechanism

- 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  $\phi$  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

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

<sup>142</sup> where the covariant derivative is given by

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

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

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

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

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

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

<sup>166</sup> Where  $H$  is a real scalar field, the *true vacuum* Higgs field. Substituting this  
<sup>167</sup> 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  $W_\mu^\pm$ ,  $Z_\mu$  and  $A_\mu$   
169 in terms of the original fields  $A_\mu^a$  and  $B_\mu$ . This gives

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

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

### 205 Lepton Masses

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

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

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  $\phi$  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  $\phi$  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  $\phi$ . 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

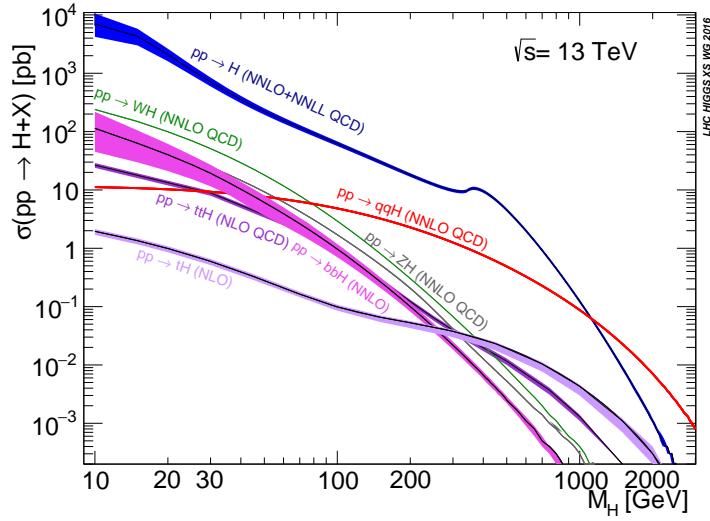
$$\begin{bmatrix} \tilde{d} \\ \tilde{s} \\ \tilde{b} \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}, \quad (1.26)$$

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

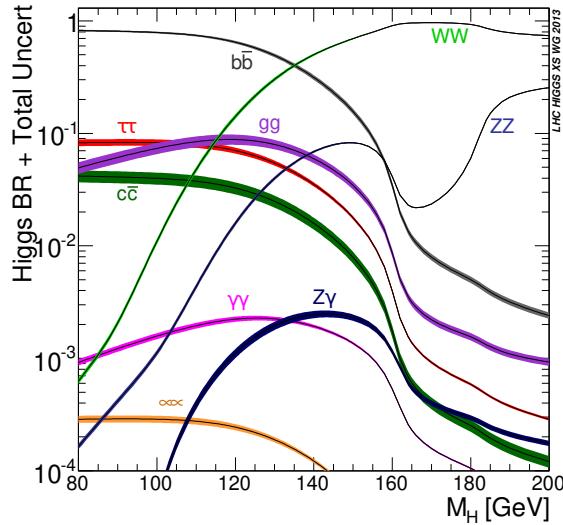
### 237 1.2.3 Higgs Phenomenology

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.

<sup>273</sup> **Chapter 2**

<sup>274</sup> **The Large Hadron Collider and the  
275 ATLAS Detector**

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

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

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	$8 \times 10^{33}$
Run 2	2015–2018	13	34	25	$1–2 \times 10^{34}$
Run 3	2022–2025	13.6	50	25	$2 \times 10^{34}$

**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 2.1.2). Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.

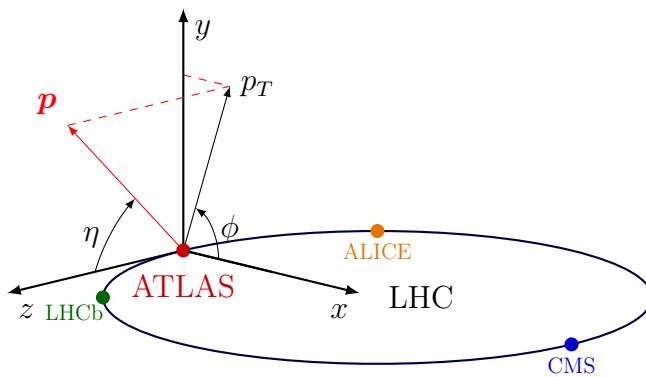
## 297 2.1 Coordinate System & Collider Definitions

### 298 2.1.1 ATLAS Coordinate System

299 ATLAS uses a right-handed coordinate system with its origin at the nominal inter-  
300 action point in the centre of the detector and the  $z$ -axis along the beam pipe. The  
301  $x$ -axis points from the interaction point to the centre of the LHC ring, and the  $y$ -axis  
302 points upwards. Cylindrical coordinates  $(r, \phi)$  are used in the transverse plane,  $\phi$   
303 being the azimuthal angle around the  $z$ -axis. The pseudorapidity is defined in terms  
304 of the polar angle  $\theta$  as  $\eta = -\ln \tan(\theta/2)$ . Angular distance is measured in units of  
305  $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ .

The transverse plane lies in  $x$ - $y$  while the longitudinal plane lies along the  $z$ -axis.

this is from pubcom



**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. Reproduced from Ref. [33].

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

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

308 Differences in  $\eta$  are invariant under Lorentz boosts. Particle production is constant  
309 as a function of  $\eta$ .

310 The transverse momentum  $p_T$  is the sum in quadrature of the momenta in the  
311 transverse plane

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

312 Transverse and longitudinal impact parameters (IP)  $d_0$  and  $z_0$  specify the closest  
313 approach of the trajectory of a particle to the origin. The transverse IP  $d_0$  and  
314 longitudinal IP  $z_0$  are measured with respect to the hard scatter primary vertex  
315 (see Section 2.3.2). Impact parameter significances are defined as the IP divided  
316 by its corresponding uncertainty,  $s(d_0) = d_0/\sigma(d_0)$  and  $s(z_0) = z_0/\sigma(z_0)$ . Track IP  
317 significances are lifetime signed according to the track's direction with respect to the  
318 jet axis and the primary vertex [34].

### 319 2.1.2 Collider Definitions

#### 320 Luminosity

321 The luminosity is defined by

xy full def

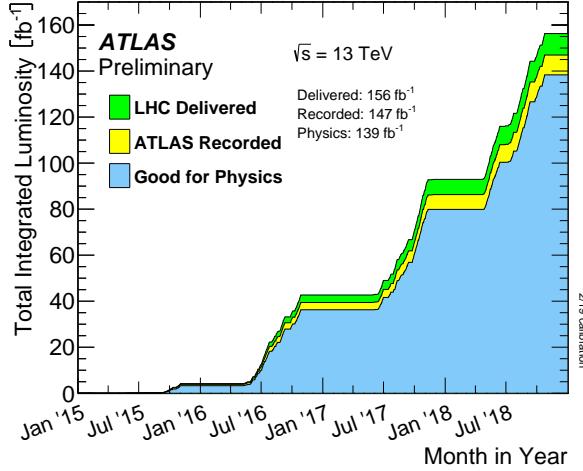
$$L = \frac{1}{\sigma} \frac{dN}{dt} = \frac{fn_b N_1 N_2}{2\pi\sigma_x\sigma_y}, \quad (2.3)$$

322 where  $N$  is the number of individual proton-proton collisions and  $\sigma$  is the cross-  
323 sectional area of the beams as they cross.

324 Total number of events is related to the total inelastic cross section

define

$$N = \sigma \int \mathcal{L} dt \quad (2.4)$$



**Figure 2.2:** Delivered, recorded, and usable integrated luminosity as a function of time during Run 2 [32].

### 325 Pile-up

326 At the centre of the ATLAS detector, bunches of more than  $10^{11}$  protons meet head  
 327 on. Each bunch-crossing is called an *event*. There is generally at most one hard  
 328 proton-proton scatter per event. Additional interactions are typically relatively soft  
 329 and are known as *pile-up*. Pile-up complicates the reconstruction of the hard scatter  
 330 event as results of the interactions of different proton-proton interactions have to  
 331 be separated. Pile-up from interactions within the same bunch-crossing is known  
 332 as *in-time* pile-up while residual signatures from other bunch-crossings is known as  
 333 *out-of-time* pile-up. The number of pile-up interactions is denoted  $\mu$ , which is often  
 334 given as a time-averaged value  $\langle \mu \rangle$ . The average number of pile-up interactions for  
 335 different years during Run 2 is given in Fig. 2.3.

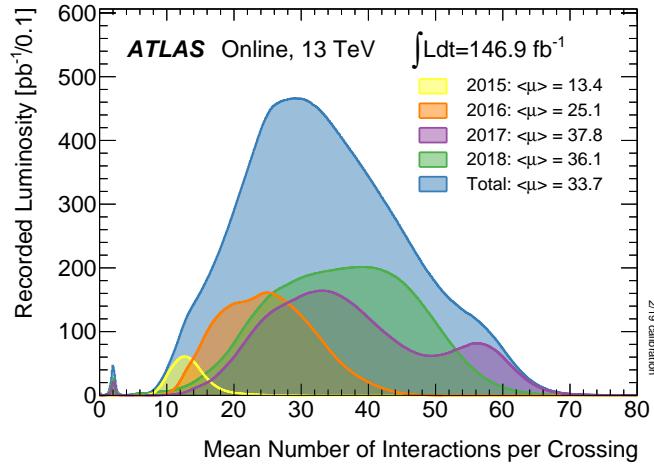
## 336 2.2 The ATLAS Detector

337  
 338 The ATLAS<sup>1</sup> detector at the LHC covers nearly the entire solid angle around the  
 339 collision point. It consists of an inner tracking detector surrounded by a thin  
 340 superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon  
 341 spectrometer incorporating three large superconducting air-core toroidal magnets.

suggestion:  
make the  
first sen-  
tence more  
of a inuti-  
tive under-  
standing,  
and then go  
into barel  
eta encap  
jargon

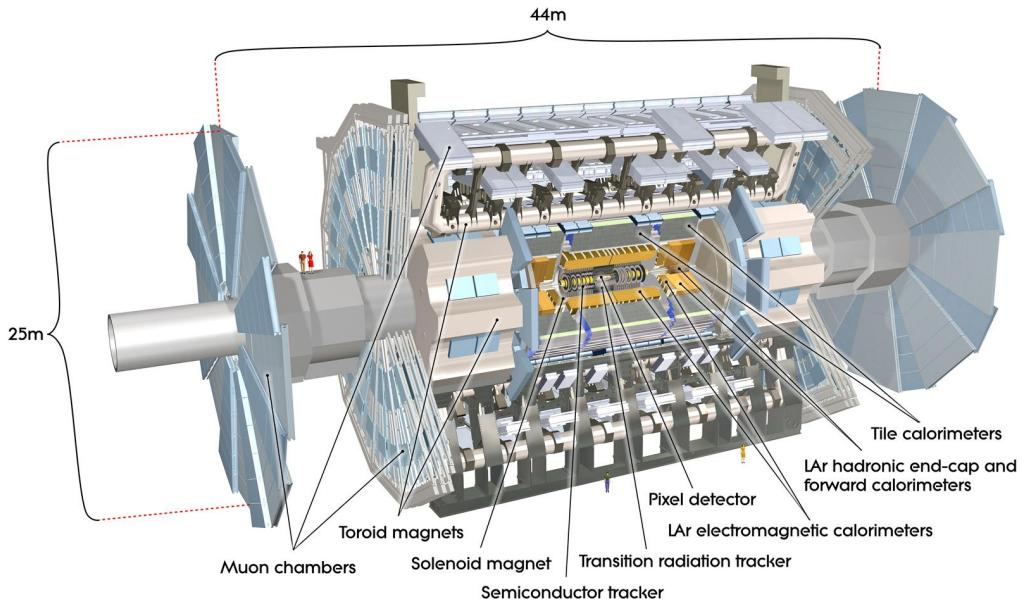
<sup>1</sup>A Toroidal Lhc ApparatuS.

okay this  
is the stan-  
dard pub-  
com text?



**Figure 2.3:** Average pile-up profiles measured by ATLAS during Run 2 [32]. During Run 3, even higher levels of pile-up are expected.

342 The detector is made up of several specialised sub-detectors as shown in Fig. 2.4. In  
 343 this section a condensed overview of each sub-detector is given, in order of increasing  
 344 radial distance from the point of collision. A more complete picture can be found in  
 345 Ref. [30], or in the technical design reports (TDRs) of the individual sub-detectors.

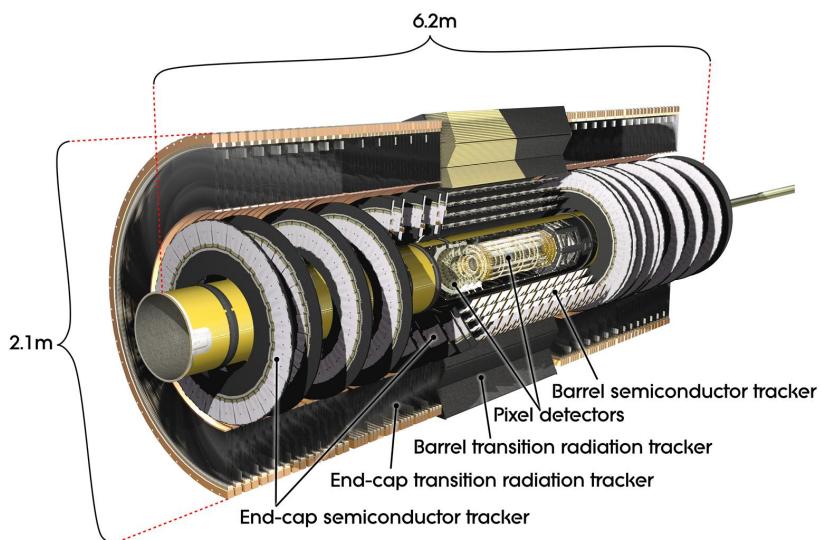


**Figure 2.4:** A 3D model of the entire ATLAS detector. Cutouts through the detector

### <sup>346</sup> 2.2.1 The Inner Detector

<sup>347</sup> The inner-detector system (ID) provides high-resolution charged particle trajectory  
<sup>348</sup> tracking in the range  $|\eta| < 2.5$ . The ID is immersed in a 2 T axial magnetic field,  
<sup>349</sup> produced by a superconducting solenoidal magnet, which enables the measurement  
<sup>350</sup> of particle transverse momentum and charge<sup>2</sup>. After Run 3, the ID will be replaced  
<sup>351</sup> by the ITk [35, 36].

<sup>352</sup> The inner detector is made up of several sub-systems, shown in Figs. 2.5 and 2.6. The  
<sup>353</sup> high-granularity silicon pixel detector covers the vertex region and typically provides  
<sup>354</sup> four spacepoint measurements per track. It is followed by the silicon microstrip  
<sup>355</sup> tracker (SCT), which usually provides a further four spacepoint measurements per  
<sup>356</sup> track. These silicon detectors are complemented by the Transition Radiation Tracker  
<sup>357</sup> (TRT), which enables radially extended track reconstruction up to  $|\eta| = 2.0$ .



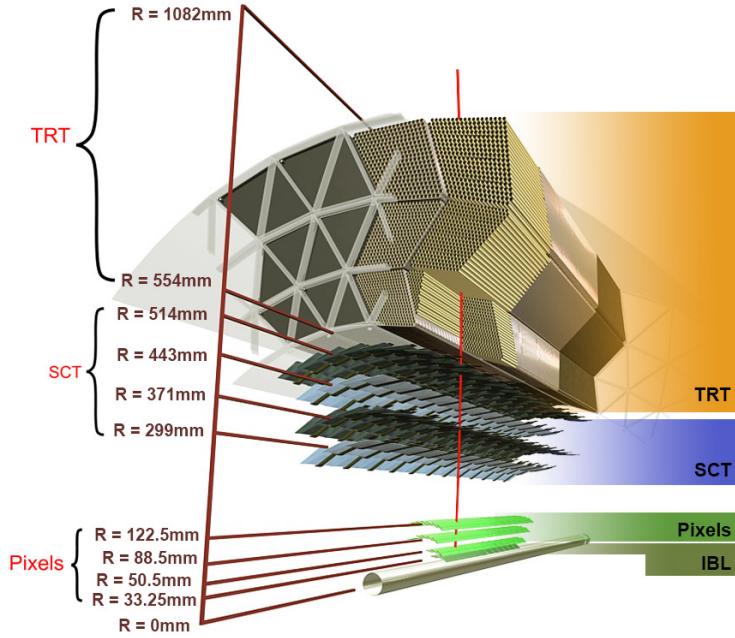
**Figure 2.5:** A 3D model of the ATLAS ID showing the barrel layers and end-cap disks.

### <sup>358</sup> Pixel Detector

<sup>359</sup> The silicon pixel detector is comprised of four cylindrical barrels at increasing radii  
<sup>360</sup> from the beamline, and four disks on each side. The innermost barrel layer is the

---

<sup>2</sup>Reconstructed charged particles are assumed to have a charge of  $\pm 1$ .

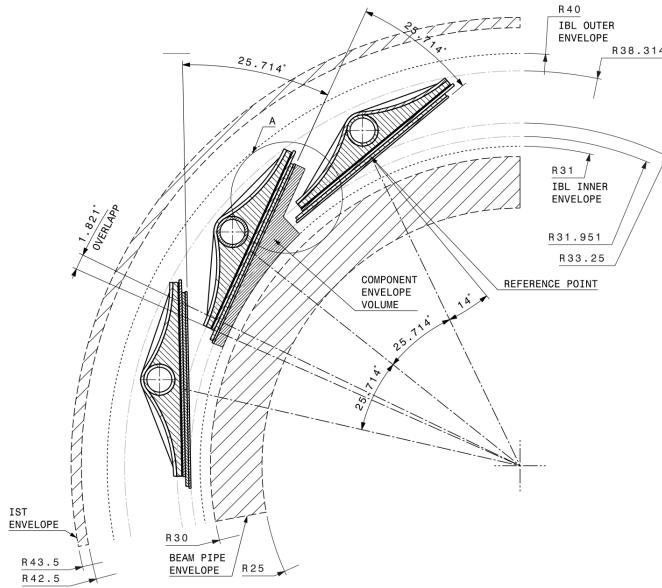


**Figure 2.6:** A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown.

361 insertable B-layer (IBL), which was installed before Run 2 [37,38] and lies just 33 mm  
 362 from the beam axis. The second-to-innermost layer is often referred to as the B-layer.  
 363 The specification of the pixel detector determines the impact parameter resolution  
 364 and the ability to reconstruct primary and secondary vertices. The detector is  
 365 required to have a high granularity (i.e. resolution) to maintain the low occupancy  
 366 required to resolve nearby particles. Individual pixels are 50  $\mu\text{m}$  in the transverse  
 367 direction  $R\phi$  and 400  $\mu\text{m}$  in the longitudinal  $z$  direction (250  $\mu\text{m}$  for the IBL). Cluster  
 368 positions have a resolution of approximately 10  $\mu\text{m}$  in  $R\phi$  and 100  $\mu\text{m}$  in  $z$ .

### 369    **Semi-Conductor Tracker (SCT)**

370 The SCT is made up of four concentric barrel layers in the central region, and nine  
 371 disks in each end-cap. Each layer is itself made of a pair of silicon microstrip layers,  
 372 with a small stereo angle (20 mrad) between the two layers enabling the  $z$ -coordinate  
 373 measurement from a pair of strip measurements. The SCT typically provides four  
 374 precision spacepoint measurements (eight strip measurements) per track in the barrel



**Figure 2.7:** A cross-sectional view of the ATLAS IBL.

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

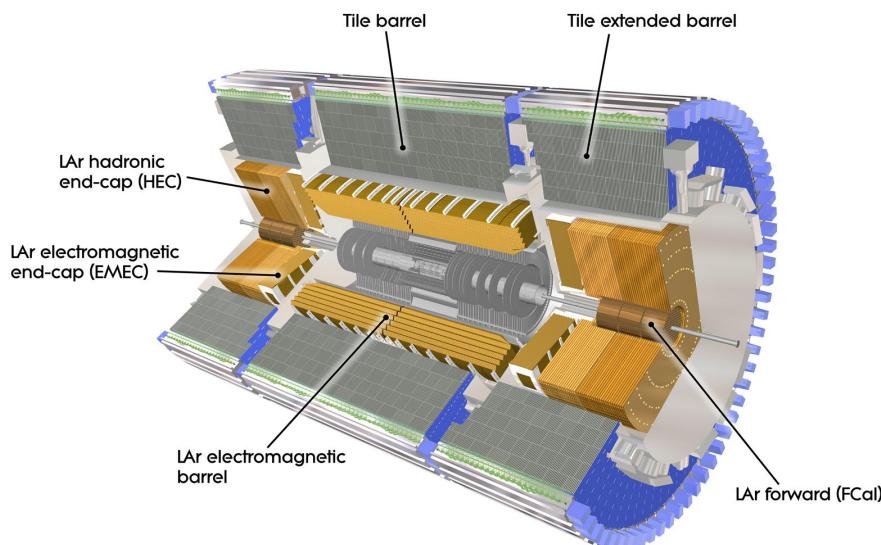
### 381 Transition Radiation Tracker (TRT)

382 The TRT is a straw-tube tracker which complements the higher-resolution silicon-based tracks by offering a larger number of hits per track (typically around 30) and 383 a long lever arm, which aids the accurate measurement of particle momentum. It is 384 made up of approximately 300 000 drift tubes with a diameter of 4 mm which are 385 filled with xenon gas. The walls of each tube are electrically charged, and a thin 386 conducting wire runs along the center. When a charged particle traverses a tube, it 387 ionises the xenon and the resulting liberated electrons drift along the electric field 388 to the wire, where an associated charge is registered. In the barrel the straws run 389 parallel to the  $z$ -axis and therefore the TRT only provides tracking information in 390

*R $\phi$ .* Straws are arranged radially in the end-caps. The resulting two-dimensional spacepoints have a resolution of approximately  $120\text{ }\mu\text{m}$ . The spaces between the straws are filled with a polymer which leads to the emission of transition radiation, aiding electron identification.

### *395* 2.2.2 Calorimeters

*396* The calorimeter system measures the energy of incident particles over the range  
*397*  $|\eta| < 4.9$ . There are two main sub-systems: the electromagnetic calorimeter (ECal),  
*398* which focuses on the measurement of electrons and photons, and the hadronic  
*399* calorimeter (HCal), which measures the energy of hadrons. Upon entering the  
*400* calorimeter, incident particles will interact with the detector material to produce a  
*401* shower of secondary particles with reduced energies. The charge deposited in this  
*402* process is measured to reconstruct the energy of the initial incident particle. The  
*403* two calorimeter sub-systems must provide strong containment of showering  
*404* particles to prevent punch-through of EM and non-muon particles to the HCal and  
*405* muon system respectively.



**Figure 2.8:** The ATLAS calorimeters. The ECal (orange) and HCal (grey, brown).

#### **406 Liquid Argon (LAr) Electromagnetic Calorimeter**

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

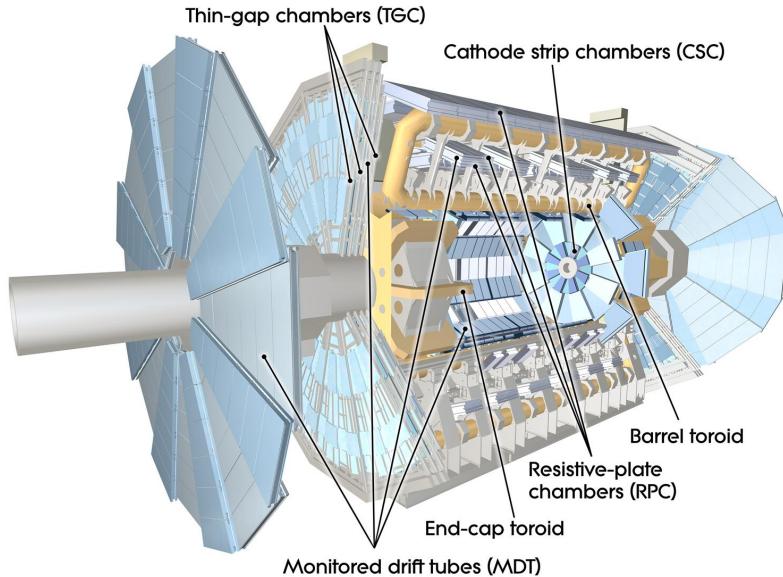
#### **415 Hadronic Tile Calorimeter**

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

### **422 2.2.3 Muon Spectrometer**

**423** Due to their higher mass, muons easily pass unimpeded through the ID and calorime-  
**424** ters and therefore require specialised detectors for their measurement. The Muon  
**425** Spectrometer (MS) is made up of dedicated tracking and triggering hardware. The  
**426** precision tracking system uses three layers of monitored drift tubes with a barrel  
**427** region covering  $|\eta| < 1.2$  and end-caps covering  $1 < |\eta| < 2.7$ . The inner layers of the  
**428** end-caps use cathode strip chambers to better cope with the high occupancy in the  
**429** forward region. Precision tracking resolution is approximately  $50\text{ }\mu\text{m}$ . The trigger  
**430** system is comprised of resistive plate chambers in the barrel region covering  $|\eta| < 1.0$   
**431** and thin gap chambers in the end-cap regions covering  $1 < |\eta| < 2.4$ . A set of three  
**432** toroidal magnets, each made up of eight coils, is used in each of the barrel and  
**433** end-caps to deflect the muons as they pass through the MS, allowing their momentum  
**434** and charge to be measured from the direction and magnitude of curvature. The

- <sup>435</sup> toroidal magnets generate a field which is largely orthogonal to the muon trajectories which allows for maximum deflection.



**Figure 2.9:** The ATLAS muon spectrometer.

<sup>436</sup>

#### <sup>437</sup> 2.2.4 The Trigger

- <sup>438</sup> The 2.5 ns bunch spacing used over the course of Run 2 corresponds to a bunch-  
<sup>439</sup> crossing or event rate of 40 MHz (see Table 2.1). If the full information for the  
<sup>440</sup> detector was written out for each event, this would correspond to the generation  
<sup>441</sup> of 60 TB of data each second. This is more than can be feasibly processed and  
<sup>442</sup> stored, requiring the use of a trigger system which quickly makes a decision about  
<sup>443</sup> whether or not an event is potentially interesting and should be kept for further  
<sup>444</sup> analysis. The trigger system is comprised of two levels which search for signs of  
<sup>445</sup> electrons, muons photons, taus and jets, as well as events with high total or missing  
<sup>446</sup> transverse energy. The hardware-based Level-1 (L1) trigger uses coarse information  
<sup>447</sup> from the calorimeters and MS to accept events at a rate of 100 kHz within 2.5  $\mu$ s  
<sup>448</sup> of the event. After the L1 trigger, the software-based High Level Trigger (HLT)  
<sup>449</sup> makes use of 40 000 CPU cores to make a final selection on surviving events within a  
<sup>450</sup> few hundred milliseconds. The final event read-out rate is approximately 1.2 kHz,

451 corresponding  $1.2 \text{ GB s}^{-1}$  of permanent data storage. More information is provided  
452 in [39].

## 453 2.3 Reconstructed Physics Objects

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

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

### 463 2.3.1 Tracks

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

### 470 Space-point Formation (Clustering)

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

## 477 Track Finding

478 Space-points are used to build track seeds. These are groups of three hits which  
479 are geometrically compatible with being part of a track segment. A combinatorial  
480 Kalman filter (KF) is used to build track candidates by extending track seeds. The  
481 filter can create multiple track candidates per seed, with bifurcations along the  
482 track occurring when more than one compatible space-point exists on a given layer.  
483 In this way, the KF creates an excessive number of *track candidates*, which are  
484 only required to satisfy basic quality requirements. Track candidates are allowed to  
485 reuse or *share* hits freely (a single hit may be used by multiple track candidates).  
486 Typically, the presence of shared hits is a predictor of a bad track due to the high  
487 granularity of the ATLAS tracking detectors. At this stage, there is a large number  
488 of incorrect hits assigned to otherwise good tracks, and additionally large number of  
489 *fake* tracks, which are comprised of a majority of wrong hits and do not correspond  
490 to the trajectory of any one physical particle (fake tracks are defined as those where  
491 the majority of associated hits do not originate from one single truth particle). The  
492 low quality of tracks at this stage necessitates an ambiguity solving step, in which  
493 candidates are cleaned, and the highest quality track are selected.

TMP def

## 494 Ambiguity Solving

495 Ambiguity solving was introduced as part of the ATLAS New Tracking effort (NEWT)  
496 [41], and is intended to improve track reconstruction performance in challenging dense  
497 environments. In the ambiguity solver, track candidates are processed individually  
498 in descending order of a track score. The track score quantifies the likelihood of the  
499 track corresponding to the trajectory of a real particle. Scoring uses a number of  
500 signals, including the number and positions of hits (preferring hits in more precise  
501 regions of the detector), the transverse momentum of the track and the track fit  
502 quality. The track fit quality describes the quality of the track as the  $\chi^2$  divided  
503 by the number degrees of freedom on the track. A preference for high transverse  
504 momentum tracks promotes the successful reconstruction of the more physically  
505 interesting energetic particles, and suppresses the large number of wrong hits assigned  
506 to low momentum tracks.

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

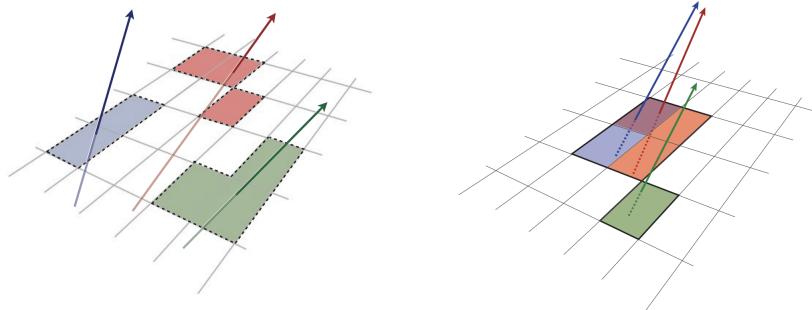
list shared  
hit cut

### 515 Neural Network Cluster Splitting

516 As part of track cleaning, shared hits are classified by a Neural Network (NN)  
517 to determine if they are compatible with the characteristic features of a merged  
518 cluster [42, 44]. A merged cluster is one which has originated from more than  
519 one incident particle. The corresponding reconstructed cluster is made up of a  
520 combination of energy deposits from more than one particle, which have become  
521 merged due to the closeness of the associated particles and the limited resolution of  
522 the detector. While in general this event is rare, it is common for clusters to become  
523 merged in dense environments, as discussed in Section 3.1. If the cluster is predicted  
524 to be merged it is labelled as being freely shareable, or *split*. Hits classified as split  
525 by the cluster splitting NN are allowed to be shared freely. Hits not compatible with  
526 the merged hypothesis can still be shared by a limited number of tracks, but come  
527 with a penalty for the track which may hinder its acceptance into the final track  
528 collection.

### 529 2.3.2 Vertices

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



**Figure 2.10:** 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.

### 536 Primary Vertices

537 Each proton-proton interaction happens at a *primary vertex* (PV). Primary vertices  
 538 are iteratively reconstructed using tracks. The *hard scatter vertex* of an event  
 539 is chosen as the primary vertex whose associated tracks have the largest sum of  
 540 transverse momentum squared,  $\Sigma p_T^2$ .

### 541 Secondary Vertices

542 Secondary vertices (SV) occur when a particle radiates or decays at a sufficient  
 543 distance from the primary vertex to be resolved from the primary vertex (see  
 544 Section 3.1.1). Two widely used secondary vertexing tools are used within ATLAS:  
 545 SV1 and JetFitter [45]. Each attempts to reconstruct secondary vertices inside a jet  
 546 given the tracks associated to that jet (see Section 2.3.3 for more information about  
 547 track association). SV1 by design attempts to reconstruct only a single inclusive  
 548 vertex per jet. This inclusive vertex groups all  $b$ -hadron decay products, including  
 549 tracks from the  $b$ -hadron decay itself and tracks from  $b \rightarrow c$  decays. The second tool,  
 550 JetFitter attempts to resolve each displaced vertex inside the jet, such that secondary  
 551 vertices from  $b$ -hadron decays are reconstructed separately to tertiary vertices from  
 552  $b \rightarrow c$  decay chains.

### 553 2.3.3 Jets

554 Jets are the reconstructed object corresponding to a spray of collimated stable  
555 particles which results from a decay chain. Jets are built by clustering constituent  
556 objects (e.g. tracks or calorimeter clusters) using a jet finding algorithm, for example  
557 the anti- $k_T$  algorithm [46].

### 558 Particle Flow Jets

559 Jets are reconstructed from particle-flow objects [47] using the anti- $k_T$  algorithm  
560 with a radius parameter of 0.4. The jet energy scale is calibrated according to  
561 Ref. [48]. Jets are also required not to overlap with a generator-level electron or  
562 muon from  $W$  boson decays. All jets are required to have a pseudorapidity  $|\eta| < 2.5$   
563 and  $p_T > 20$  GeV. Additionally, a standard selection using the Jet Vertex Tagger  
564 (JVT) algorithm at the tight working point is applied to jets with  $p_T < 60$  GeV and  
565  $|\eta| < 2.4$  in order to suppress pile-up contamination [49]. Tracks are associated  
566 to jets using a  $\Delta R$  association cone, the width of which decreases as a function of  
567 jet  $p_T$ , with a maximum cone size of  $\Delta R \approx 0.45$  for jets with  $p_T = 20$  GeV and  
568 minimum cone size of  $\Delta R \approx 0.25$  for jets with  $p_T > 200$  GeV. If a track is within the  
569 association cones of more than one jet, it is assigned to the jet which has a smaller  
570  $\Delta R(\text{track}, \text{jet})$ .

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

### 576 Track Jets

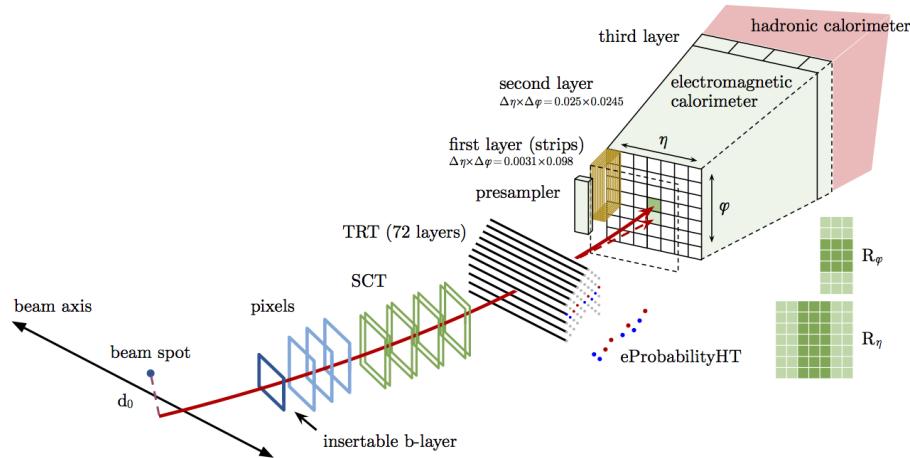
577 ???

### 578 2.3.4 Leptons

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

#### 582 Electrons

583 Electrons candidates are reconstructed by matching PV-compatible inner detector  
 584 tracks to calorimeter clusters. The track-cluster matching criteria takes into account  
 585 the significant energy loss of the electron due to bremsstrahlung. If a match is  
 586 found, a refit of the track is performed using the Gaussian Sum Filter (GSF) [50],  
 587 which better handles trajectory reconstruction in the presence of bremsstrahlung.  
 588 Various identification criteria are then applied using a likelihood-based (LH) method  
 589 to the candidates to improve purity. These include requirements on the track quality  
 590 and cluster matching, the shape of shower in the ECal, leakage into the HCal, and  
 591 the amount of transition radiation detected in the TRT. A full description can be  
 obtained in Ref. [51].



**Figure 2.11:** A sketch of electron reconstruction using the ATLAS detector [51]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.

**593 Muons**

594 Muon reconstruction primarily makes use of the dedicated MS (see Section 2.2.3), but  
595 also relies on tracks from the ID and the presence of characteristic signatures in the  
596 calorimeters. Muon tracks are reconstructed by connect straight-line track segments,  
597 which are identified via a Hough transform, and combined into a approximately  
598 parabolic trajectory. Finally, a global  $\chi^2$  fit is performed, taking into account possible  
599 interactions between the muon and the detector material. A reconstructed muon is  
600 called *combined* if it completes successful matching to an ID track. Combined muons  
601 undergo a further fit with the combined ID and MS hits, with the energy loss due to  
602 the traversal of the calorimeters being taking into account.

603 **Chapter 3**

604 **Tracking and  $b$ -tagging**

605 Many ATLAS analyses rely on a method of tagging jets instantiated by  $b$ -quarks  
606 ( $b$ -jets) and rejecting jets created from other quarks ( $c$  and light flavours  $u$ ,  $d$ ,  $s$ ).  
607 These  $b$ -tagging algorithms work by discriminating for the unique signatures of  $b$ -jets  
608 discussed in Section 3.1.  $b$ -tagging relies on the efficient and accurate reconstruction  
609 the tracks corresponding to the  $b$ -hadron decay products, as these tracks are used as  
610 the primary inputs to vertex reconstruction algorithms, jet making algorithms and  
611 jet tagging algorithms.

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

616 **3.1  $b$ -hadron Reconstruction**

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

619 **3.1.1 Decay Topology**

620  $b$ -hadrons are quasi-stable bound states of a bottom quark and one or more lighter  
621 quarks. Collectively, these are the  $b$ -mesons (e.g.  $B^+ = u\bar{b}$ ,  $B^0 = d\bar{b}$ ) and baryons

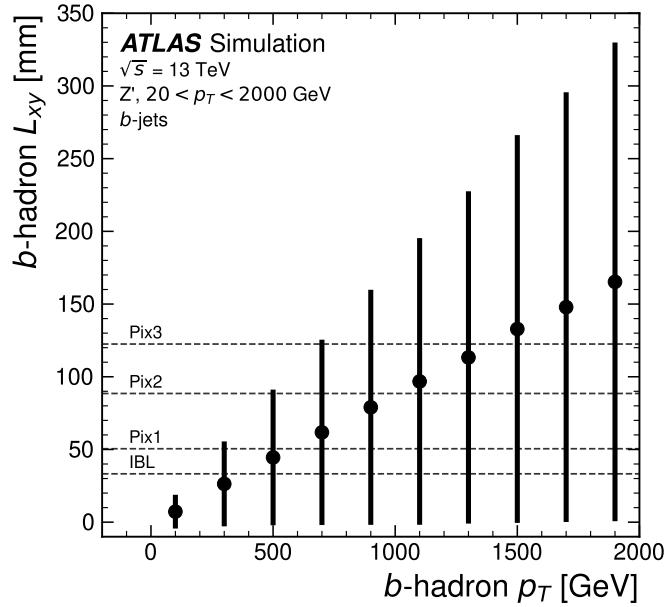
with (e.g.  $\Lambda_b^0 = udb$ ). After a *b*-quark is produced as the result of some proton-proton collision, they quickly hadronise. The hadronisation process is hard – around 70-80% of the *b*-quark’s momentum is passed to the *b*-hadron, with the rest being radiated as prompt hadronisation or fragmentation particles. See Ref. [52] for a more in depth discussion on hadronisation and the closely related process of fragmentation. Henceforth the combined hadronisation and fragmentation products will be referred to collectively as fragmentation.

*b*-hadrons are interesting objects of study due to their relatively long proper life-times  $\tau \approx 1.5$  ps. Early studies showed that *b*-hadrons did not couple strongly to light-flavour quarks [53]. The lifetime of *b*-hadrons is therefore approximately determined only by a single CKM matrix element  $V_{cb}$  (see Section 1.1.3). This lifetime corresponds to a proper decay length  $c\tau \approx 450$   $\mu\text{m}$ . In the rest frame of the detector, the typical *b*-hadron travels a distance

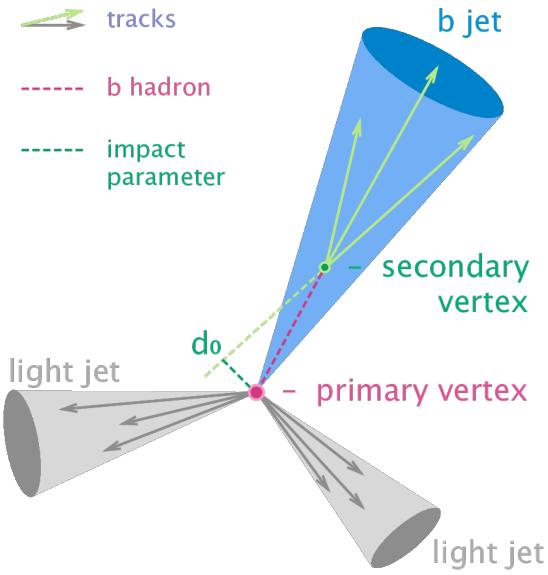
$$d = \gamma\beta c\tau \approx \gamma c\tau \quad (3.1)$$

before decaying, where in the high energy limit  $\gamma = E_b/m_b$  and  $\beta = v/c = 1$ . For a 1 TeV *b*-hadron, this gives  $d \approx 90$  mm – well beyond the radius of the first pixel layer (the IBL) at a radius of 33 mm (see Fig. 3.1). This significant displacement is characteristic of *b*-jets and makes it possible to reconstruct secondary vertices at the *b*-hadron decay point.

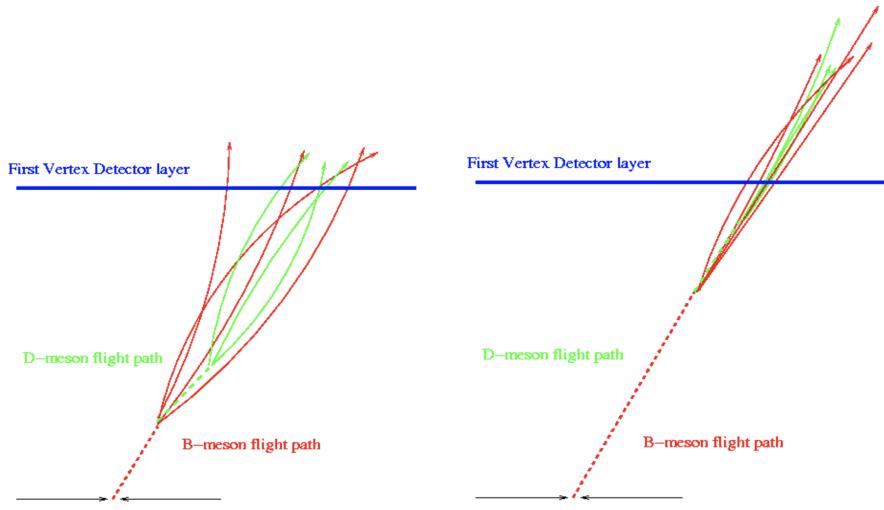
*b*-hadrons decay weakly to on average four or five collimated stable particles. These particles, along with any fragmentation particles, are reconstructed in the detector as a jet. A *b*-jet has several characteristic features which differentiate it from light-jets. These features stem from the significant 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. 3.2). 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.



**Figure 3.1:** The truth  $b$ -hadron decay radius  $L_{xy}$  as a function of truth transverse momentum  $p_T$  for reconstructed  $b$ -jets in  $Z'$  events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines.



**Figure 3.2:** 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.

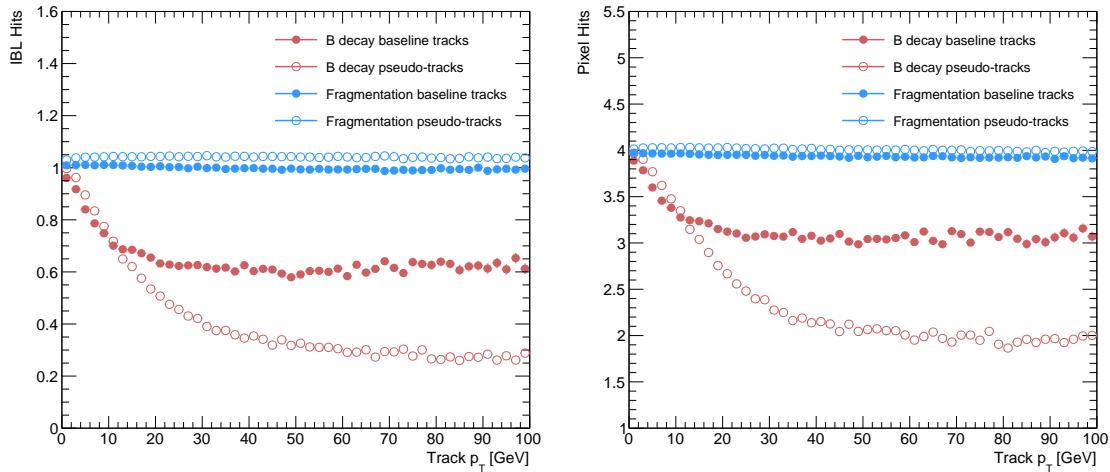


**Figure 3.3:** At lower  $p_T$  (left) the decay length of the  $b$ -hadron is reduced, and the resulting decay tracks are less collimated. At higher  $p_T$  (right) the  $b$ -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge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

### 650 3.1.2 Challenges Facing $b$ -hadron Reconstruction

651 As discussed, a necessary requirement for successful jet  $b$ -tagging is the efficient  
 652 and accurate reconstruction of the charged particle trajectories in the jet. For high  
 653  $p_T$  jets ( $p_T > 200$  GeV) this task becomes difficult due to a combination of effects.  
 654 As the jet energy increases, the multiplicity of tracks in the jet increases due to  
 655 the presence of additional fragmentation particles. Fragmentation and weak decay  
 656 products also become increasingly collimated as their inherited transverse momentum  
 657 increases. Together, these two effects lead to a high density of charged particles in  
 658 the jet core, which, given the finite resolution of the detector, makes reconstruction  
 659 difficult. At high energies, the increased decay length of  $b$ -hadrons (and  $c$ -hadrons)  
 660 means that decay products have less of an opportunity to diverge before reaching  
 661 the first tracking layers of the detector (shown in Fig. 3.3). If the weak decay takes  
 662 place close enough to a detector layer, or if the particles are otherwise sufficiently  
 663 collimated, charge deposits left by nearby particles may not be resolved individually,  
 664 instead being reconstructed as merged clusters. As discussed in Section 2.3.1, merged  
 665 clusters are generally rare, and so shared hits generally predict bad tracks and  
 666 are correspondingly penalised during track reconstruction. However, in the core of

high  $p_T$   $b$ -jets the density of particles is high enough that the probability of cluster merging increases dramatically. Successful reconstruction of such tracks requires the presence of shared hits, but the presence of these can paradoxically end up impairing the successfully reconstruction of the track. Furthermore, decays may also take place inside the tracking detectors themselves, which at best leads to missing measurements on the most sensitive detector layers, and at worst can lead to wrong inner layer hits being added to displaced tracks, since the reconstruction process penalises tracks without inner layer hits. The combination of effects described above 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.

The above effects create two related, but distinct problems for  $b$ -tagging. The first part is a drop in track reconstruction efficiency. As mentioned, tracks originating from high energy  $b$ -hadron decay products can have a high rate of shared hits due to the number of particles present in a high  $p_T$   $b$ -jet and their relative collimation. Additionally, tracks may be missing hits on the inner layers of the detector in the

681 case of displaced decays. The presence of shared and missing hits reduces a track's  
 682 score in the ambiguity solver meaning that higher ranking, but potentially worse,  
 683 track candidates are processed first and take ownership of the hits. This can make  
 684 it difficult for otherwise reasonable  $b$ -hadron decay tracks to meet the ambiguity  
 685 solver's stringent track quality requirements, leading to their rejection at this stage.  
 686 this is shown in Fig. 3.5 .

get a nice  
reco eff plot

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

699 The combination of reduced reconstruction efficiency and incorrectly assigned hits is  
 700 thought to be the cause of the observed drop in  $b$ -tagging efficiency at high energies,  
 701 however it is not clear which effect may dominate.

$b$ -tagging vs  
pT plot

## 702 3.2 Investigating Improvements for High $p_T$ B 703 Tracking

704 An investigation into

### 705 3.2.1 Pseudotrack and Ideal Tracks

706 Pseudo-tracking uses Monte Carlo truth information to group together all the hits left  
 707 by each truth particle. Each group of hits which passes basic quality requirements is  
 708 directly used in a full resolution track fit. If the track fit is successful, a “pseudo-track”

709 track is created and stored. If the track fit fails, or the collection of hits does not pass  
 710 the basic quality requirements (for example because of a lack of hits) then the particle  
 711 is said to be un-reconstructable. In this way, pseudo-tracking performance represents  
 712 the ideal reconstruction performance given the ATLAS detector, with perfect hit-  
 713 to-track association and and track reconstruction efficiency. The approach was  
 714 introduced in [54] as a way to obtain a fast approximation of tracking reconstruction  
 715 for simulated data, however the technique has become a useful tool for studying  
 716 tracking performance in general [42].

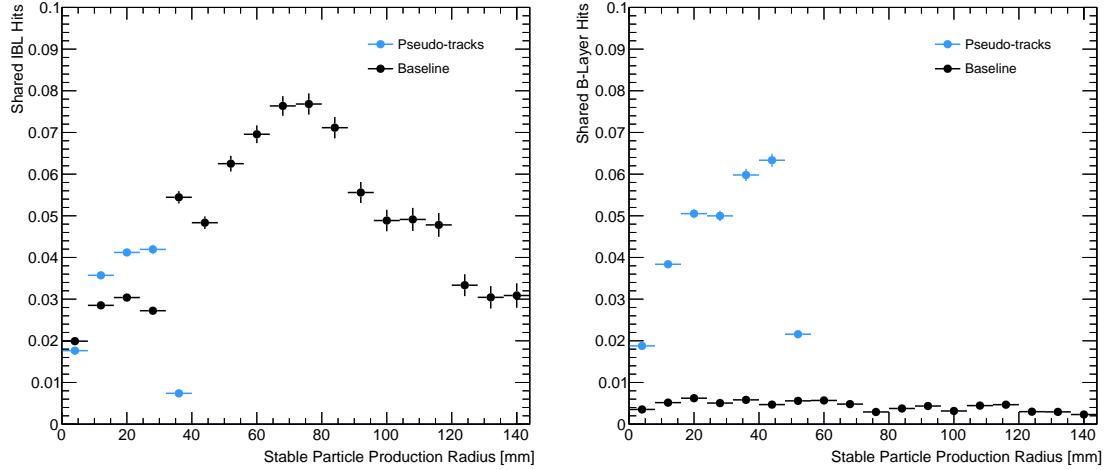
717 The ambiguity solver is not run for pseudo-tracks. However, if the standard track  
 718 collection is produced alongside the pseudo-tracks, then cluster splitting neural  
 719 networks will be run for the standard tracks, and the resulting classification of  
 720 clusters will be propagated to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

### 733 3.2.2 Global $\chi^2$ Fitter Outlier Removal

734

735 This section documents ongoing progress into improving hit assignments using the  
 736 Global  $\chi^2$  Fitter (GX2F) to prevent wrong hits from being assigned to tracks during  
 737 the track fit. This is in contrast to the approach discussed in cref:sec:refit, which  
 738 attempts to identify and remove wrong hits after the reconstruction of the track (of  
 739 which the track fit is a part). As part of the track fit, an outlier removal procedure  
 740 is run, in which suspicious hits are identified and removed. The GX2F code, as a  
 741 relatively low-level component of track reconstruction, has not undergone significant

discuss  
whether to  
keep this  
section



**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.

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  $\chi^2$  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 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 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.

---

## 760 Implementation

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

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

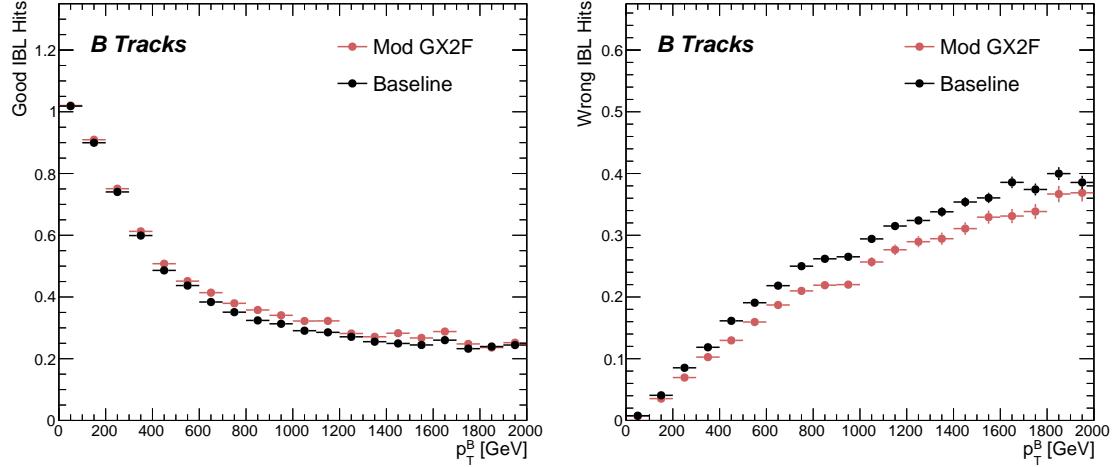
770 This pull is computed for the transverse and longitudinal coordinates of the mea-  
 771 surement, and the maximum of the two is selected and checked to see if it exceeds a  
 772 certain threshold. If it does, the hit will be removed, after some additional checks are  
 773 made to confirm or deny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

## 775 Cut Optimisation

776 A systematic variation of the cut point `m_outlcut` has been carried out. The value of  
 777 `m_outlcut` was reduced from 4 down to 1.75, a change which affects a silicon layers  
 778 (the TRT has separate outlier removal logic). Furthermore, a specific cut for the IBL  
 779 was introduced, and is set to 1.25. A second cut, `TrackChi2PerNDFCut`, is also used  
 780 in the outlier removal. This value was reduced from 7 to 4. Finally, instead of taking  
 781 the maximum of the pulls in the longitudinal and transverse directions, a quadrature  
 782 sum is taken of these two values and used. This variation is labelled “Mod GX2F” in  
 783 plots.

784 The results, demonstrating a reduction in wrong hit assignment whilst also improving  
 785 slightly the good good hits assigned to tracks, are shown in Fig. 3.7. The improve-  
 786 ments are also observed when looking inclusively in all tracks, which removes the  
 787 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



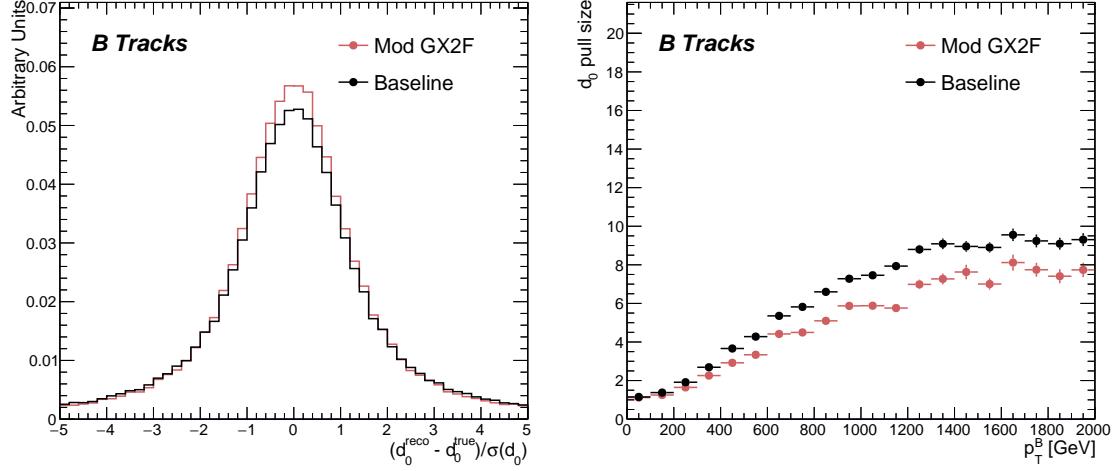
**Figure 3.7:** Profiles, as a function of parent  $b$ -hadron  $p_T$ , of good (left) and wrong (right) hit assignment rates on the IBL for tracks using baseline tracking (black), the modified version of the outlier removal procedure (red).

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

## Conclusion

The ambiguity solving process relies on many pre-defined cuts which have not been optimised for high transverse momentum  $b$ -hadron track reconstruction. A full investigation

Though the results show some improvement over baseline tracking, the potential enhancement of these changes to  $b$ -tagging would need to be assessed before putting them in production. As there are some known data-MC discrepancies, fine tuned optimisation such as the work presented here presents an opportunity to over-optimize the tracking algorithms to MC.



**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.

### 805 3.3 Tracking software validation

- 806     • tracking validation
- 807     • qspi validation

808 **Chapter 4**

809 **Track Classification MVA**

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

819 **4.1 Machine Learning Background**

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

829 programmed rules can fail to fully capture the relationships between inputs and  
 830 outputs.

831 In the simplest case, a set of  $m$  labelled training examples  $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$   
 832 is collected. Each element  $(x_i, y_i)$  consists of a input vector  $x_i \in \mathbb{R}^{\text{input}}$ , and the  
 833 corresponding label  $y_i$ . In classification problems, these labels are integer *class labels*  
 834  $y_i \in \{0, \dots, N - 1\}$ , where  $N$  is the number of classes, which specify which of a  
 835 pre-determined set of categorical classes the training example belongs to. The rest  
 836 of the discussion in this chapter is limited to binary classification problems ( $N = 2$ ).  
 837 The two classes are often referred to as signal ( $y_i = 1$ ) and background ( $y_i = 0$ ),  
 838 which need to be separated. Collecting sufficient and suitable data is one of the  
 839 primary challenges of machine learning, as such data is not always readily available.  
 840 Fortunately, sophisticated tools to simulate particle collisions have already been  
 841 developed by the scientific community [55, 56]. Simulation plays a key role in particle  
 842 physics. More detail is given in.

843 After obtaining suitable training data, the next step is to define a model. Given an  
 844 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  
 845 parameterised functional mapping from input space to output space. Given an input  
 846 example  $x_i$  and a set of parameters  $\theta$ , the model outputs a prediction  $\hat{y}_i \in (0, 1)$  for  
 847 the true label  $y_i$ , as in

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

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

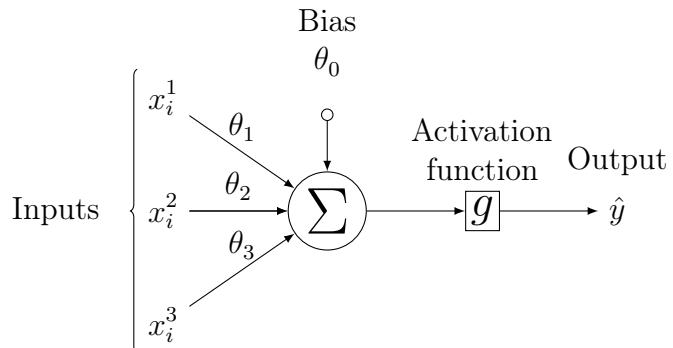
add section  
somewhere  
on sample  
info

### 855 4.1.1 Neural Networks

856 Neural networks (NNs) are a common choice for the machine learning model  $f$  since  
 857 they have the ability to approximate any function [57] and are easy to train via  
 858 backpropagation [58].

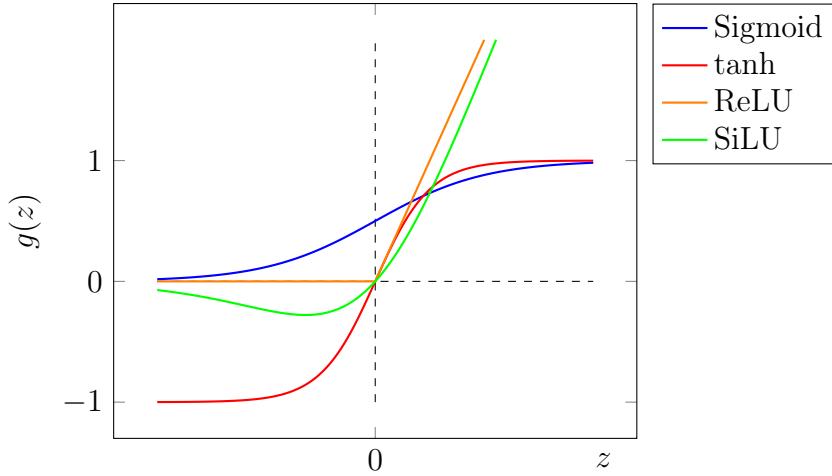
### 859 Artificial Neurons

860 The basic functional component of a NN is the *artificial neuron* or node, which is  
 861 loosely inspired by a mathematical model of a biological neuron [59, 60]. An artificial  
 862 neuron is defined by its parameters or *weights*  $\theta$  and a choice of activation function.  
 863 Each neuron takes a fixed number of inputs and computes the dot product of the  
 864 input and weight vectors  $x^T \theta$  and additionally adds a constant bias term  $\theta_0$ . This  
 865 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  $\theta_j$ , and the resulting values are summed. A bias term  $\theta_0$  is added, and the result is passed to an activation function. Each neuron can be thought of as a logistic regression model.

866 The dot product is fed into an activation function  $g$ . The activation function has  
 867 several uses, most notably acting as a source of non-linearity and bounding the  
 868 output of the neuron. Some common activation functions are shown in Fig. 4.2.  
 869 The choice of activation function can have implications for the performance and  
 870 convergence of the network, since the gradient of  $g$  is used to compute the weight  
 871 updates during training. This is also why input data is normalised to have zero mean  
 872 and unity variance [61].



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

873 **Networks**

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

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

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

884 **4.1.2 Training with Gradient Descent**

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

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

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

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

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

895 The hyperparameter  $\alpha$  is known as the *learning rate* and dictates the size of the  
 896 step taken in the direction of the slope. The errors for each parameter are efficiently  
 897 calculated using the backpropagation algorithm. The process of updating weights  
 898 is repeated until the weights converge the network is trained. In practice, small  
 899 batches of the input data are shown to the network at a time. For each batch the  
 900 average loss is calculated and the network's weights are updated. There are many  
 901 extensions and variations of the gradient descent algorithm. This work uses the Adam  
 902 optimiser which adds momentum to the weight updates (dampening oscillations)  
 903 and an adaptive per-parameter learning rate [62].

## 904 4.2 Track Truth Origin Labelling

905 Crucial to supervised learning techniques are the ground truth class labels which  
 906 the machine learning model is trained to predict. A set of track truth labels which a  
 907 high degree of granularity have been implemented, and are listed in Table 4.1. For the

908 fake track classification tool, the origins in Table 4.1 are used to construct a binary  
 909 label by labelling all fake tracks which are not also fromB as background, and all other  
 910 tracks (i.e. good tracks and fake fromB tracks) as signal. The fake track classifier is  
 911 then trained to distinguish between these two categories of tracks. Fake tracks are  
 912 defined using the *truth-matching probability* (TMP), defined in Eq. (4.5). This is a  
 913 weighted sum of the number of hits on a track which are from the same truth particle,  
 914 versus the total number of hits on the track. The weights are subdetector-dependent

---

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  $\Lambda^0$  decays, and hadronic interactions.

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

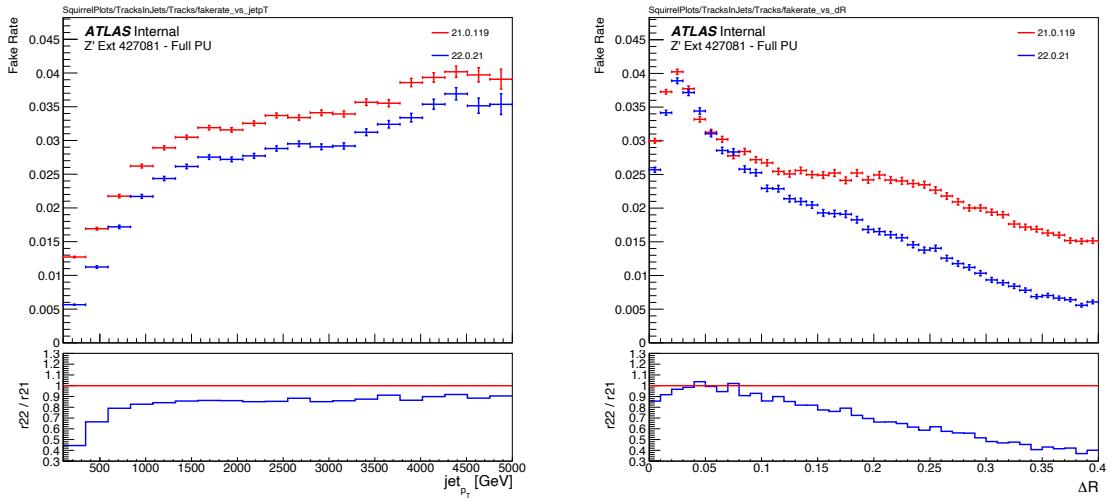
$$\text{TMP} = \frac{10N_{\text{Pix}}^{\text{good}} + 5N_{\text{SCT}}^{\text{good}} + N_{\text{TRT}}^{\text{good}}}{10N_{\text{Pix}}^{\text{all}} + 5N_{\text{SCT}}^{\text{all}} + N_{\text{TRT}}^{\text{all}}} \quad (4.5)$$

916 The labelling scheme has been designed to be useful beyond the classification of good and  
917 fake tracks.

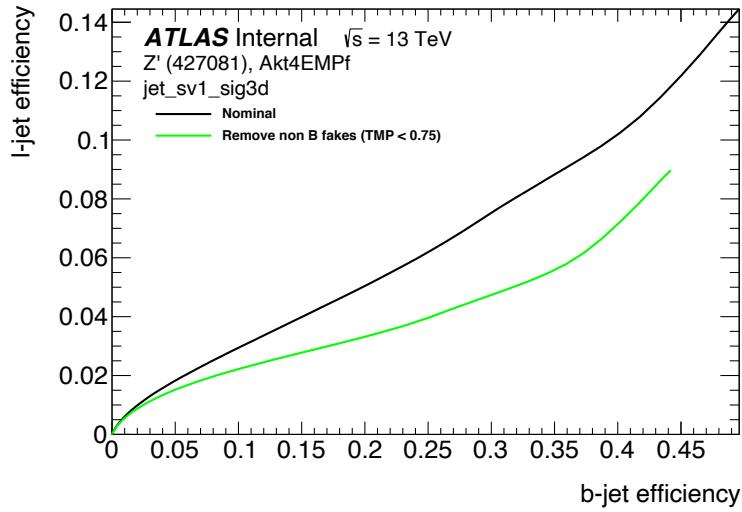
### 918 4.3 Fake Track Identification Tool

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

923 To identify and remove fake tracks, a NN classification tool was trained with good  
924 tracks as the signal class and fake tracks as the background class. Due to the  
925 imbalance between the two classes (with fake tracks being relatively uncommon),  
926 a weight was added to the loss function for the background class to account for  
927 this. The NN was made up of two hidden layers with 100 nodes each. The ReLU  
928 activation function was used in conjunction with the Adam optimiser.



**Figure 4.3:** Rate of fake tracks as a function of jet transverse momentum (left) and  $\Delta R$ (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).



**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.

929 Inputs to the model are described in Section 4.3.1, while fake track removal performance  
930 is given in Section 4.3.2.

### 931 4.3.1 Model Inputs

932 The fake track MVA is given two jet variables and 20 tracking related variables  
933 for each track fed into the network. The jet transverse momentum and signed  
934 pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in  
935 Table 4.2. Full track parameter information and associated uncertainties, along with  
936 detailed hit information, carry valuable information about the jet flavour.

copied from  
GN1, but  
fake track  
mva uses  
deta instead  
of deta dphi

937 Track selection follows the loose selection described in Ref. [63] and outlined in  
938 Table 4.3, which was found to improve the flavour tagging performance compared to  
939 previous tighter selections, whilst ensuring good resolution of tracks and a low fake  
940 rate [43].

- 941 • network structure
- 942 • trained on x tracks
- 943 • train test
- 944 • input variables
- 945 • samples – borrow from gnn note
- 946 • performance plots
- 947 • impact on btagging

### 948 4.3.2 Results

## 949 4.4 Conclusion

950 Fake tracks, which are prevalent in the core of high  $p_T$  jets, are shown to have an  
951 adverse impact on  $b$ -tagging performance. A ML tool to identify fake tracks has been

Jet Input	Description
$p_T$	Jet transverse momentum
$\eta$	Signed jet pseudorapidity
Track Input	Description
$p_T$	Track transverse momentum
$\Delta R$	Angular distance between the track and jet
$d_0$	Closest distance from the track to the PV in the longitudinal plane
$z_0$	Closest distance from the track to the PV in the transverse plane
nIBLHits	Number of IBL hits
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nTRTHits	Number of TRT hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixOutliers	Number of pixel outliers
nSCTOutliers	Number of SCT outliers
nTRTOutliers	Number of TRT outliers
$r$	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 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.

Parameter	Selection
$p_T$	$> 500$ MeV
$ d_0 $	$< 3.5$ mm
$ z_0 \sin \theta $	$< 5$ mm
Silicon hits	$\geq 8$
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  $\theta$  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.

MVA Output Cut	Good Track Efficiency		Fake Track Efficiency	
	All	FromB	All	FromB
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.

952 developed, which can be used to limit the number of fake tracks being inputted to  
953 downstream  $b$ -tagging algorithms. These fake or poorly reconstructed tracks could  
954 also be remove in other places 

keep?

955 Another application of this tool is to isolate a relatively pure sample of fake tracks to  
956 estimate the fake track rate in data, which can be used for tracking recommendations.

957 A more advanced approach than detecting and removing fake tracks is to attempt  
958 to determine whether a track is good or fake, but allow any downstream tagging  
959 algorithms to use this information when coming to a descision about wether to tag a  
960 jet, rather than removing the track as an input.

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

963 Preliminary studies into a version of the tool which are used to select heavy flavour  
964 tracks from  $b$ -hadron decays have have led to the development of a mutliclassification  
965 approach...

966 **Chapter 5**

967 **Graph Neural Network Flavour  
968 Tagger**

969 Flavour tagging, the identification of jets originating from  $b$ - and  $c$ -quarks, is a  
970 critical component of the physics programme of the ATLAS experiment at the Large  
971 Hadron Collider. Current flavour tagging algorithms rely on the outputs of several  
972 low-level algorithms, which reconstruct various properties of jets using charged  
973 particle tracks, that are then combined using machine learning techniques. In this  
974 note a new machine learning algorithm based on graph neural networks, GN1, is  
975 introduced. GN1 uses information from a variable number of charged particle tracks  
976 within a jet, to predict the jet flavour without the need for intermediate low-level  
977 algorithms. Alongside the jet flavour prediction, the model predicts which physics  
978 processes produced the different tracks in the jet, and groups tracks in the jet into  
979 vertices. These auxiliary training objectives provide useful additional information on  
980 the contents of the jet and improve performance. GN1 compares favourably with  
981 the current ATLAS flavour tagging algorithms. For a  $b$ -jet efficiency of 70%, the  
982 light ( $c$ )-jet rejection is improved by a factor of  $\sim 1.8$  ( $\sim 2.1$ ) for jets coming from  
983  $t\bar{t}$  decays with transverse momentum  $20 < p_T < 250$  GeV. For jets coming from  $Z'$   
984 decays with transverse momentum  $250 < p_T < 5000$  GeV, the light ( $c$ )-jet rejection  
985 improves by a factor  $\sim 6$  ( $\sim 2.8$ ) for a comparative 30%  $b$ -jet efficiency.

986 Work in this chapter has been published [64]. Figures, tables and text from the  
987 published note are reproduced here.

## 988 5.1 Motivation

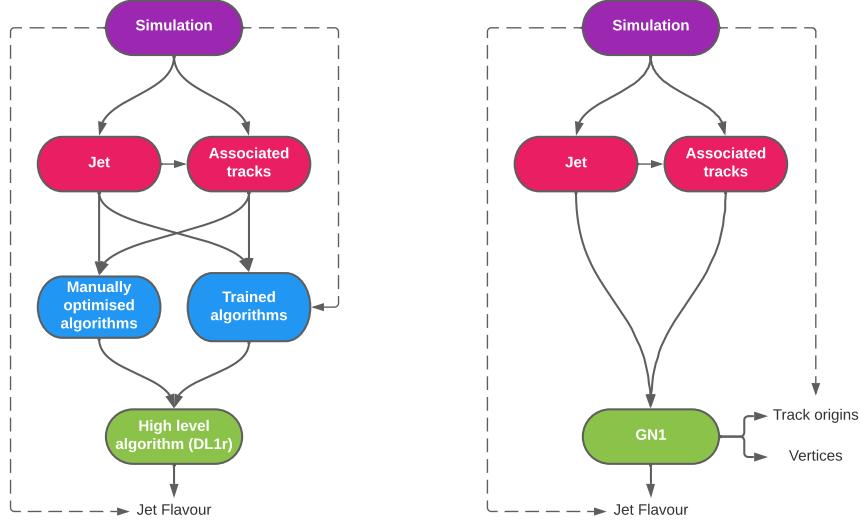
989 Flavour tagging, the identification of jets originating from  $b$ - and  $c$ -quarks, is a  
990 critical component of the physics programme of the ATLAS experiment [30] at the  
991 Large Hadron Collider (LHC) [65]. It is of particular importance for the study of the  
992 Standard Model (SM) Higgs boson and the top quark, which preferentially decay  
993 to  $b$ -quarks [66, 67], and additionally for several Beyond Standard Model (BSM)  
994 resonances that readily decay to heavy flavour quarks [68]. The significant lifetime  
995 of  $b$ -hadrons, approximately 1.5 ps [69], provides the unique signature of a secondary  
996 decay vertex which has a high mass and is significantly displaced from the primary  
997 vertex. Additional signatures of  $b$ -hadrons are the tertiary decay vertex, resulting  
998 from  $b \rightarrow c$  decay chains, and the reconstructed trajectories of charged particles  
999 (henceforth simply referred to as tracks) with large impact parameters<sup>1</sup> (IPs). These  
1000 signatures are primarily identified using tracks associated to jets. As such, efficient  
1001 and accurate track reconstruction is essential for high performance flavour tagging.

1002 This note introduces a novel algorithm, GN1, which uses Graph Neural Networks  
1003 (GNNs) [70] with auxiliary training objectives, to aid the primary goal of classifying  
1004 whether jets originate from  $b$ - or  $c$ -quarks (referred to as a flavour tagger). The  
1005 concept is illustrated in Fig. 5.1. The use of GNNs offers a natural way to classify  
1006 jets with variable numbers of unordered associated tracks, while allowing for the  
1007 inclusion of auxiliary training objectives [71, 72].

1008 The current ATLAS flavour tagger, DL1r [73], is a deep neural network which takes  
1009 the outputs of a number of independently optimised “low-level” algorithms [45] as  
1010 inputs. Each of these low-level algorithms makes use of tracks to reconstruct a  
1011 particular aspect of the experimental signature of heavy flavour jets. The low-level  
1012 algorithms can be manually optimised reconstruction algorithms, for example the  
1013 SV1 and JetFitter algorithms that reconstruct displaced decay vertices, or trained  
1014 taggers such as RNNIP and DIPS that use the IPs of a variable number of tracks  
1015 to identify the flavour of the jet [45, 63, 74, 75]. In contrast GN1 utilises a single  
1016 neural network, which directly takes the tracks and some information about the jet  
1017 as inputs. As such, it does not depend on any other flavour tagging algorithm, and a  
1018 single training of the GN1 fully optimises all aspects of the algorithm.

---

<sup>1</sup>The 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.

1019 GN1 is trained to understand the internal structure of the jet through the use of  
1020 two auxiliary training objectives: the grouping of tracks originating from a common  
1021 vertex, and the prediction of the underlying physics process from which each track  
1022 originated. These auxiliary objectives are meant to guide the neural network towards  
1023 a more complete understanding of the underlying physics, removing the need for the  
1024 low-level algorithms, and therefore simplifying the process of optimising the tagger for  
1025 new regions of phase space (e.g.  $c$ -tagging or high- $p_T$   $b$ -tagging), or when the detector  
1026 or charged particle reconstruction algorithms are updated. The training targets  
1027 for the primary and auxiliary objectives are extracted from “truth information”, i.e.  
1028 information only available in simulation, as opposed to reconstructed quantities  
1029 available in both collision data and simulation.

1030 In this note, the following benefits of this approach will be shown:

- 1031 1. Improved performance with respect to the current ATLAS flavour tagging  
1032 algorithms, with larger background rejection for a given signal efficiency.
- 1033 2. The same network architecture can be easily optimised for a wider variety of use  
1034 cases (e.g.  $c$ -jet tagging and high- $p_T$  jet tagging), since there are no low-level  
1035 algorithms to retune.
- 1036 3. There are fewer flavour tagging algorithms to maintain.
- 1037 4. Alongside the network’s prediction of the jet flavour, the auxiliary vertex and  
1038 track origin predictions provide more information on why a jet was (mis)tagged  
1039 or not. This information can also have uses in other applications, for instance  
1040 to explicitly reconstruct displaced decay vertices or to remove fake tracks.<sup>2</sup>

1041 This note is organised as follows: a brief description of the ATLAS detector, object  
1042 definitions and selections, and samples are provided in Section 5.3; details about the  
1043 model architecture and training procedure are given in Section 5.4; and results are  
1044 discussed in Section 5.5.

---

<sup>2</sup>A 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].

## 5.2 Graph Neural Network Theory

## 5.3 Experimental Setup

### 5.3.1 Datasets

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

The  $t\bar{t}$  events are generated using the POWHEGBox [76–79] v2 generator at next-to-leading order with the NNPDF3.0NLO [80] set of parton distribution functions (PDFs). The  $h_{\text{damp}}$  parameter<sup>3</sup> is set to 1.5 times the mass of the top-quark ( $m_{\text{top}}$ ) [81], with  $m_{\text{top}} = 172.5$  GeV. The events are interfaced to PYTHIA 8.230 [82] to model the parton shower, hadronisation, and underlying event, with parameters set according to the A14 tune [83] and using the NNPDF2.3LO set of PDFs [84].  $Z'$  events are generated with PYTHIA 8.2.12 with the same tune and PDF set. The decays of  $b$ - and  $c$ -hadrons are performed by EVTGEN v1.6.0 [85]. Particles are passed through the ATLAS detector simulation [86] based on GEANT4 [87].

For the  $t\bar{t}$  events, at least one  $W$  boson from the top quark decay is required to decay leptonically. Truth labelled  $b$ -,  $c$ - and light- jets are kinematically re-sampled in  $p_T$  and  $\eta$  to ensure identical distributions in these variables. The resulting dataset contains 30 million jets, 60% of which are  $t\bar{t}$  jets and 40% of which are  $Z'$  jets. While DL1r uses 70%  $t\bar{t}$  jets and 30%  $Z'$  jets, the change in sample composition did not affect the final performance of GN1. To evaluate the performance of the model, 500k jets from both the  $t\bar{t}$  and  $Z'$  samples, which are statistically independent from the

---

<sup>3</sup>The  $h_{\text{damp}}$  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.

1072 training sample, are used. Track- and jet-level inputs are scaled to have a central  
1073 value of zero and a variance of unity before training and evaluation.

## 1074 5.4 Model Architecture

### 1075 5.4.1 Model Inputs

1076 GN1 is given two jet variables and 21 tracking related variables for each track fed into  
1077 the network. The jet transverse momentum and signed pseudorapidity constitute the  
1078 jet-level inputs, with the track-level inputs listed in Table 5.1. If a jet has more than 40  
1079 associated tracks, the first 40 tracks with the largest transverse IP significance<sup>4</sup>  $s(d_0)$   
1080 are selected as inputs. Full track parameter information and associated uncertainties,  
1081 along with detailed hit information, carry valuable information about the jet flavour.  
1082 In the dense cores of high- $p_T$  jets, tracks are highly collimated and separation between  
1083 tracks can be of the same order as the active sensor dimensions, resulting in merged  
1084 clusters and tracks which share hits [43]. Due to the relatively long lifetimes of  
1085  $b$ -hadrons and  $c$ -hadrons, which can traverse several layers of the ID before decaying  
1086 and have highly collimated decay products, the presence of shared or missing hits is  
1087 a critical signature of heavy flavour jets.

1088 Dependence on the absolute value of the azimuthal jet angle  $\phi$  is explicitly removed  
1089 by providing only the azimuthal angle of tracks relative to the jet axis. The track  
1090 pseudorapidity is also provided relative to the jet axis.

1091 Since heavy flavour hadrons can decay semileptonically, the presence of a recon-  
1092 structed lepton in the jet carries discriminating information about the jet flavour.  
1093 In addition to the baseline GN1 model, the GN1 Lep variant includes an additional  
1094 track-level input, leptonID, which indicates if the track was used in the reconstruction  
1095 of an electron, a muon or neither. The muons are required to be combined [88],  
1096 and the electrons are required to pass the *VeryLoose* likelihood-based identification  
1097 working point [89].

---

<sup>4</sup>Impact 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
$\eta$	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 $\eta$
$d\phi$	Azimuthal angle of the track, relative to the jet $\phi$
$d_0$	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin \theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on $q/p$
$\sigma(\theta)$	Uncertainty on track polar angle $\theta$
$\sigma(\phi)$	Uncertainty on track azimuthal angle $\phi$
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used to reconstruct an electron or muon

**Table 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.

1098 Track selection follows the loose selection described in Ref. [63] and outlined in Ta  
1099 ble 5.2, which was found to improve the flavour tagging performance compared to  
1100 previous tighter selections, whilst ensuring good resolution of tracks and a low fake  
1101 rate [43].

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

**Table 5.2:** Quality selections applied to tracks, where  $d_0$  is the transverse IP of the track,  $z_0$  is the longitudinal IP with respect to the PV and  $\theta$  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.

#### 1102 5.4.2 Auxiliary Training Objectives

1103 In addition to the jet flavour classification, two auxiliary training objectives are  
1104 defined. Each auxiliary training objective comes with a training target which, similar  
1105 to the jet flavour label, are truth labels derived from the simulation. The presence  
1106 of the auxiliary training objectives improves the jet classification performance as  
1107 demonstrated in Section 5.5.3.

1108 The first auxiliary objective is the prediction of the origin of each track within the  
1109 jet. Each track is labelled with one of the exclusive categories defined in Table 4.1  
1110 after analysing the particle interaction that led to its formation. Since the presence  
1111 of different track origins is strongly related to the flavour of the jet, training GN1 to  
1112 recognise the origin of the tracks may provide an additional handle on the classification  
1113 of the jet flavour. This task may also aid the jet flavour prediction by acting as a

1114 form of supervised attention [90] - in detecting tracks from heavy flavour decays the  
1115 model may learn to pay more attention to these tracks.

1116 Displaced decays of  $b$ - and  $c$ -hadrons lead to secondary and tertiary vertices inside  
1117 the jet. Displaced secondary vertices can also occur in light-jets as a result of material  
1118 interactions and long-lived particle decays (e.g.  $K_S^0$  and  $\Lambda^0$ ). The second auxiliary  
1119 objective is the prediction of track-pair vertex compatibility. For each pair of tracks  
1120 in the jet, GN1 predicts a binary label, which is given a value 1 if the two tracks  
1121 in the pair originated from the same point in space, and 0 otherwise. To derive the  
1122 corresponding truth labels for training, truth production vertices within 0.1 mm are  
1123 merged. Track-pairs where one or both of the tracks in the pair have an origin label  
1124 of either Pileup or Fake are given a label of 0. Using the pairwise predictions from  
1125 the model, collections of commonly compatible tracks can be grouped into vertices.  
1126 The addition of this auxiliary training objective removes the need for inputs from a  
1127 dedicated secondary vertexing algorithm.

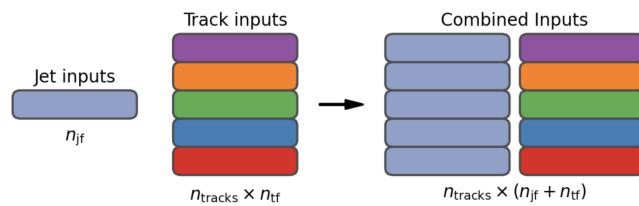
1128 Both auxiliary training objectives can be considered as “stepping stones” on the way  
1129 to classifying the flavour of the jet. By requiring the model to predict the truth  
1130 origin of each track and the vertex compatibility of each track-pair, the model is  
1131 guided to learn representations of the jet which are connected to the underlying  
1132 physics and therefore relevant for classifying the jet flavour.

### 1133 5.4.3 Architecture

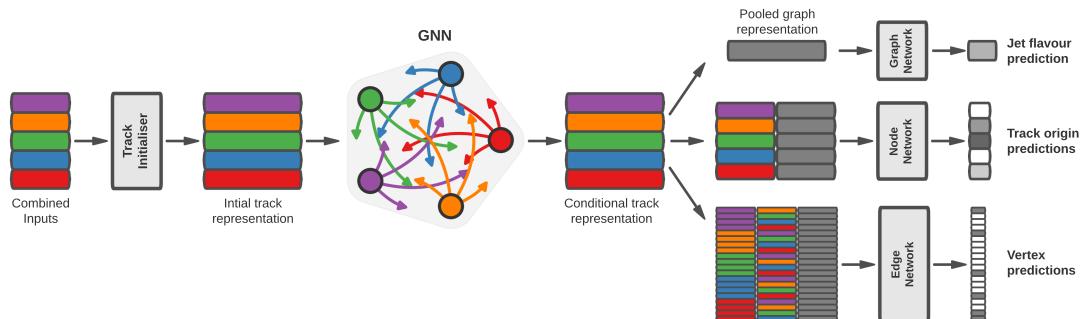
1134 As discussed above, the GN1 model combines a graph neural network architecture [91]  
1135 with auxiliary training objectives in order to determine the jet flavour. Coarse  
1136 optimisation of the network architecture hyperparameters, for example number of  
1137 layers and number of neurons per layer, has been carried out to maximise the tagging  
1138 efficiency.

1139 The model architecture is based on a previous implementation of a graph neural  
1140 network jet tagger [72]. As compared to the previous approach, GN1 uses a only a  
1141 single graph neural network and makes use of a more sophisticated graph neural net-  
1142 work layer [92], described below. These changes yield improved tagging performance  
1143 and a significant reduction in training time with respect to the previous approach.

1144 The model takes jet- and track-level information as inputs, as detailed in Section 5.4.1.  
 1145 The jet inputs are concatenated with each track’s inputs, as shown in Fig. 5.2. The  
 1146 combined jet-track vectors are then fed into a per-track initialisation network with  
 1147 three hidden layers, each containing 64 neurons, and an output layer with a size of  
 1148 64, as shown in Fig. 5.3. The track initialisation network is similar to a Deep Sets  
 1149 model [93], but does not include a reduction operation (mean or summation) over  
 1150 the output track representations.



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



**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.

1151 A fully connected graph is built from the outputs of the track initialisation network,  
 1152 such that each node in the graph neighbours every other node. Each node  $h_i$  in  
 1153 the graph corresponds to a single track in the jet, and is characterised by a feature  
 1154 vector, or representation. The per-track output representations from the initialisation  
 1155 networks are used to populate the initial feature vectors of each node in the graph.

1156 In each layer of the graph network, output node representations  $h'_i$  are computed by  
1157 aggregating the features of  $h_i$  and neighbouring nodes  $\mathcal{N}_i$  as described in Ref. [92].  
1158 First, the feature vectors of each node are fed into a fully connected layer  $\mathbf{W}$ , to  
1159 produce an updated representation of each node  $\mathbf{W}h_i$ . These updated feature vectors  
1160 are used to compute edge scores  $e(h_i, h_j)$  for each node pair,

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

1161 where  $\oplus$  denotes vector concatenation,  $\theta$  is a non-linear activation function, and  $\mathbf{a}$  is  
1162 a learned vector. These edge scores are then used to calculate attention weights  $a_{ij}$   
1163 for each pair of nodes using the softmax function over the edge scores

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

1164 Finally, the updated node representation  $h'_i$  is computed by taking the weighted sum  
1165 over each updated node representation  $\mathbf{W}h_i$ , with weights  $a_{ij}$

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

1166 The above set of operations constitute a single graph network layer. Three such layers  
1167 are stacked to construct the graph network, representing a balance between achieving  
1168 optimal performance and preventing overtraining. The final output node feature  
1169 vectors from the network are representations of each track that are conditional on the  
1170 other tracks in the jet. The output representation for each track is combined using a  
1171 weighted sum to construct a global representation of the jet, where the attention  
1172 weights for the sum are learned during training. Three separate fully connected  
1173 feedforward neural networks are then used to independently perform the different  
1174 classification objectives of GN1. Each of the objectives makes use of the global

1175 representation of the jet. A summary of the different classification networks used for  
1176 the various training objectives is shown in Table 5.3.

**Table 5.3:** 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

1177 A node classification network, which takes as inputs the features from a single output  
1178 node from the graph network and the global jet representation, predicts the track  
1179 truth origin, as defined in Table 4.1. This network has three hidden layers containing  
1180 128, 64 and 32 neurons respectively, and an output size of seven, corresponding to  
1181 the seven different truth origins.

1182 An edge classification network, which takes as inputs the concatenated representations  
1183 from each pair of tracks and the global jet representation, is used to predict whether  
1184 the tracks in the track-pair belong to a common vertex. The edge network has three  
1185 hidden layers containing 128, 64 and 32 neurons respectively, and a single output,  
1186 which is used to perform binary classification of the track-pair compatibility. These  
1187 predictions are used for the auxiliary training objectives discussed in Section 5.4.2.

1188 A graph classification network takes only the global jet representation as an input,  
1189 and predicts the jet flavour. The graph classification network is comprised of four  
1190 fully connected hidden layers with 128, 64, 32 and 16 neurons respectively, and has  
1191 three outputs corresponding to the  $b$ -,  $c$ - and light- jet classes.

#### 1192 5.4.4 Training

1193 The full GN1 training procedure minimises the total loss function  $L_{\text{total}}$ , defined in  
1194 Eq. (5.4). This loss is composed of three terms:  $L_{\text{jet}}$ , the categorical cross entropy  
1195 loss over the different jet flavours;  $L_{\text{vertex}}$ , the binary track-pair compatibility cross  
1196 entropy loss averaged over all track-pairs; and  $L_{\text{track}}$ , the categorical cross entropy loss

1197 for the track origin prediction.  $L_{\text{vertex}}$  is computed by averaging over all track-pairs  
1198 in the batch, and  $L_{\text{track}}$  is computed by averaging over all tracks in the batch.

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

1199 The different losses converge to different values during training, reflective of differences  
1200 in the relative difficulty of the various objectives. As such,  $L_{\text{vertex}}$  and  $L_{\text{track}}$  are  
1201 weighted by  $\alpha = 1.5$  and  $\beta = 0.5$  respectively to ensure they converge to similar values,  
1202 giving them an equal weighting towards  $L_{\text{total}}$ . The values of  $\alpha$  and  $\beta$  also ensure  
1203 that  $L_{\text{jet}}$  converges to a larger value than  $L_{\text{vertex}}$  and  $L_{\text{track}}$ , reflecting the primary  
1204 importance of the jet classification objective. In practice, the final performance of  
1205 the model was not sensitive to modest variations in the loss weights  $\alpha$  and  $\beta$ , or to  
1206 pre-training using  $L_{\text{total}}$  and fine tuning on the jet classification task only. As there  
1207 was a significant variation in the relative frequency of tracks of different origins, the  
1208 contribution of each origin class to  $L_{\text{track}}$  was weighted by the inverse of the frequency  
1209 of their occurrence. In  $L_{\text{vertex}}$ , the relative class weight in the loss for track-pairs  
1210 where both tracks are from either a  $b$ - or  $c$ -hadron is increased by a factor of two as  
1211 compared with other track-pairs.

1212 The track classification and vertexing objectives are supplementary to the jet clas-  
1213 sification objective and trainings can be performed with either the node or edge  
1214 networks, or both, removed, as discussed in Section 5.5.3. In these cases, the cor-  
1215 responding losses  $L_{\text{vertex}}$  and  $L_{\text{track}}$  are removed from the calculation of  $L_{\text{total}}$ . The  
1216 resulting trainings demonstrate how useful the different auxiliary training objectives  
1217 are for the primary jet classification objective.

1218 GN1 trainings are run for 100 epochs on 4 NVIDIA V100 GPUs, taking around 25  
1219 mins to complete each epoch over the training sample of 30 million jets described  
1220 in Section 5.3.1. The Adam optimiser [94] with an initial learning rate of 1e−3,  
1221 and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically the  
1222 validation loss, calculated on 500k jets, stabilised after around 60 epochs. The epoch  
1223 that minimized the validation loss was used for evaluation. GN1 has been integrated  
1224 into the ATLAS software [40] using ONNX [95], and jet flavour predictions for the  
1225 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  $20 < p_T < 250 \text{ GeV}$  from the  $t\bar{t}$  sample and jets with  $250 < p_T < 5000 \text{ GeV}$  from the  $Z'$  sample. Performance is compared to the DL1r tagger [73], which has been retrained on 75 million jets from the same samples as GN1. The input RNNIP tagger [75] 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

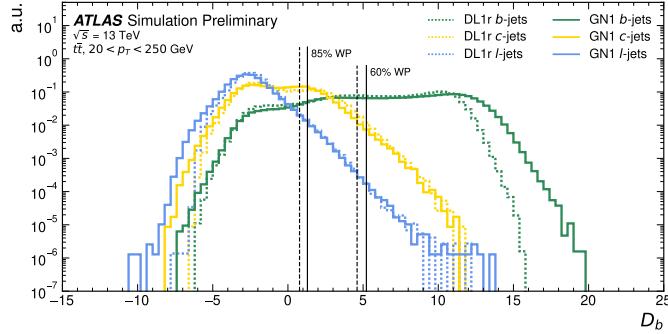
$$D_b = \log \frac{p_b}{(1 - f_c)p_l + f_c p_c}, \quad (5.5)$$

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 [73]. 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.

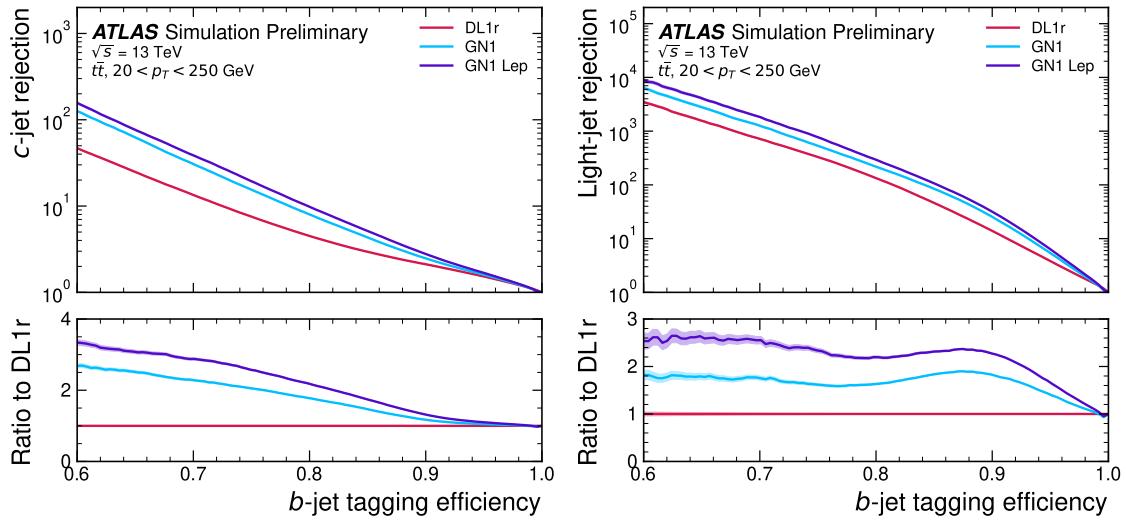
### 5.5.1 $b$ -tagging Performance

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.

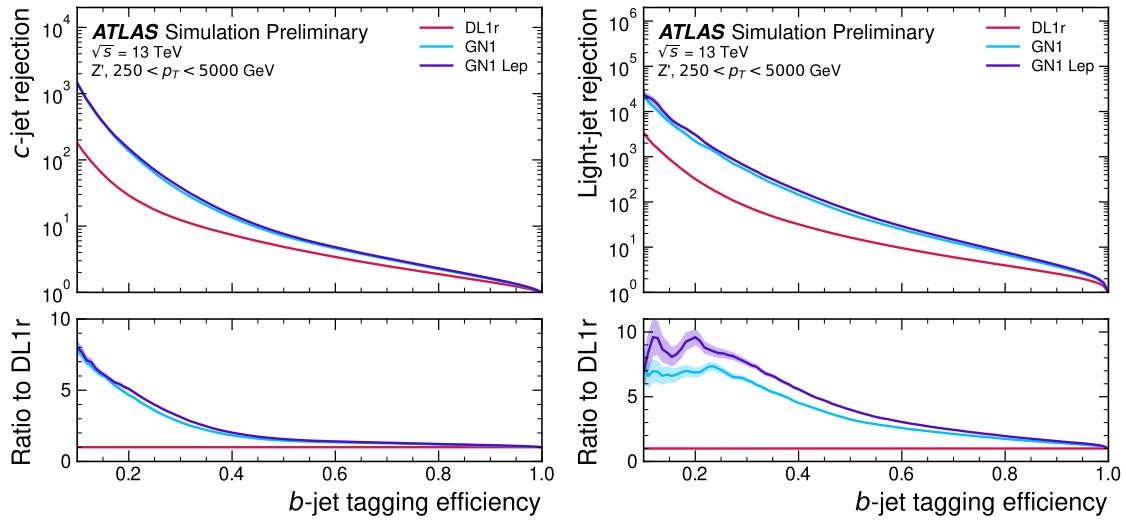
1272 For jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV, GN1 demonstrates considerably  
1273 better  $c$ - and light-jet rejection compared with DL1r across the full range of  $b$ -jet  
1274 tagging efficiencies probed. The relative improvement depends on the  $b$ -jet tagging  
1275 efficiency, with the largest improvements found at lower values. At a  $b$ -jet tagging  
1276 efficiency of 70%, the  $c$ -rejection improves by a factor of  $\sim 2.1$  and the light-jet  
1277 rejection improves by a factor of  $\sim 1.8$  with respect to DL1r. For high- $p_T$  jets in the  
1278  $Z'$  sample with  $250 < p_T < 5000$  GeV, GN1 also brings considerable performance  
1279 improvements with respect to DL1r across the range of  $b$ -jet tagging efficiencies  
1280 studied. Again, the largest relative improvement in performance comes at lower  
1281  $b$ -jet tagging efficiencies. At a  $b$ -jet tagging efficiency of 30%, GN1 improves the  
1282  $c$ -rejection by a factor of  $\sim 2.8$  and the light-jet rejection by a factor of  $\sim 6$ . An  
1283 increasing statistical uncertainty due to the high rejection of background affects the  
1284 comparison at lower  $b$ -jet tagging efficiencies. It is estimated that for a  $b$ -jet tagging  
1285 efficiency of 70% in the  $t\bar{t}$  sample,  $\sim 5\%$  ( $\sim 30\%$ ) of the relative improvement in the  
1286  $c$ -jet (light-jet) rejection comes from loosening the track selection and for a  $b$ -jet  
1287 tagging efficiency of 30% in the  $Z'$  the corresponding number is  $\sim 10\%$  for both  $c$ -jets  
1288 and light-jets. Given the sophisticated exploitation of low-level information, further  
1289 studies are needed to confirm if the performance gain is also observed in experimental  
1290 data.

1291 The GN1 Lep variant shows improved performance with respect to the baseline GN1  
1292 model, demonstrating the additional jet flavour discrimination power provided by the  
1293 leptonID track input. For jets in the  $t\bar{t}$  sample, the relative  $c$ -rejection improvement  
1294 with respect to DL1r at the 70%  $b$ -jet WP increases from a factor of  $\sim 2.1$  for GN1 to  
1295 a factor of  $\sim 2.8$  for GN1 Lep. The improvement in light-jet rejection also increases  
1296 from a factor of  $\sim 1.8$  to  $\sim 2.5$  at this WP. For jets in the  $Z'$  sample, the relative  
1297  $c$ -rejection (light-jet rejection) improvement with respect to DL1r increases from a  
1298 factor of  $\sim 2.8$  to  $\sim 3$  ( $\sim 6$  to  $\sim 7.5$ ) at a  $b$ -jet tagging efficiency of 30%. As shown in  
1299 Fig. 5.7, the greatest improvement of GN1 Lep over GN1 is seen at low  $p_T$ .

1300 The performance of the taggers is strongly dependent on the jet  $p_T$ . Charged particle  
1301 reconstruction is particularly challenging within high- $p_T$  jets [43]. The multiplicity of  
1302 fragmentation particles increases as a function of  $p_T$ , while the number of particles  
1303 from heavy flavour decays stays constant. Collimation of particles inside the jet  
1304 increases and approaches the granularity of the tracking detectors, making it difficult  
1305 to resolve the trajectories of different particles. Furthermore, at high  $p_T$ , heavy



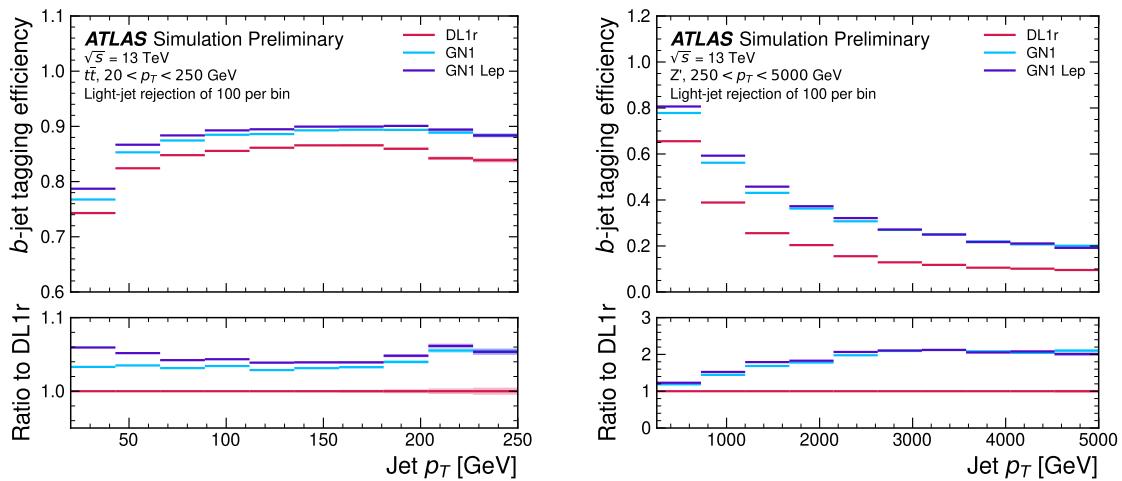
**Figure 5.5:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$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

flavour hadrons will travel further into the detector before decaying. For hadrons which traverse one or more layers of the ID before decaying, the corresponding decay tracks may pick up incorrect hits, left by the hadron itself or fragmentation particles, in the inner layers of the detector, reducing the accuracy of the reconstructed track parameters. These factors contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a general degradation in quality of tracks inside the core of a jet, which in turn reduces the jet classification performance.

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



**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.

### 5.5.2 $c$ -tagging Performance

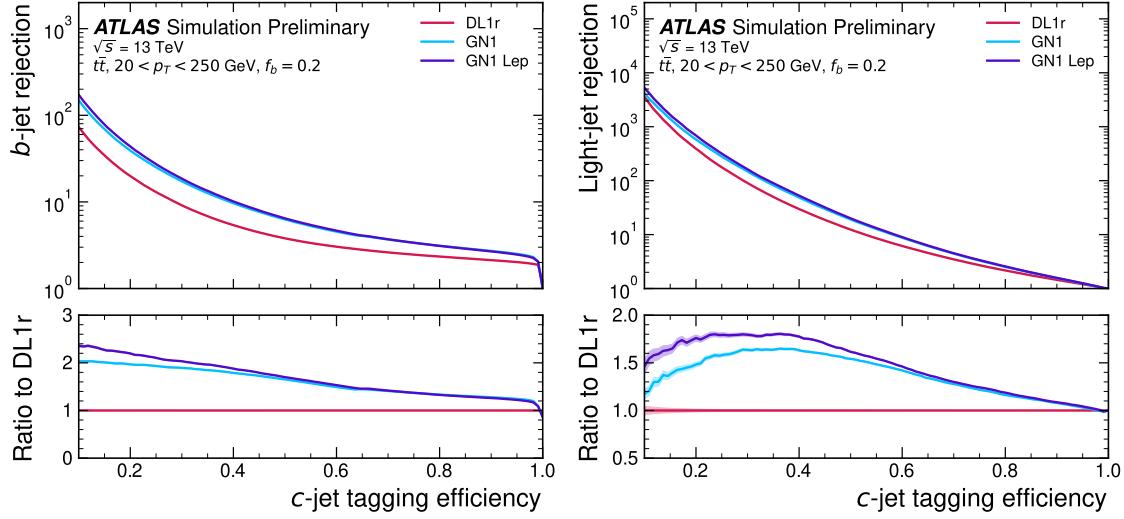
Since GN1 does not rely on any manually optimised low-level tagging algorithms, which may not have been optimised for  $c$ -tagging, tagging  $c$ -jets presents a compelling use case for GN1. To use the model for  $c$ -tagging, the output probabilities are combined into a single score  $D_c$ , defined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

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

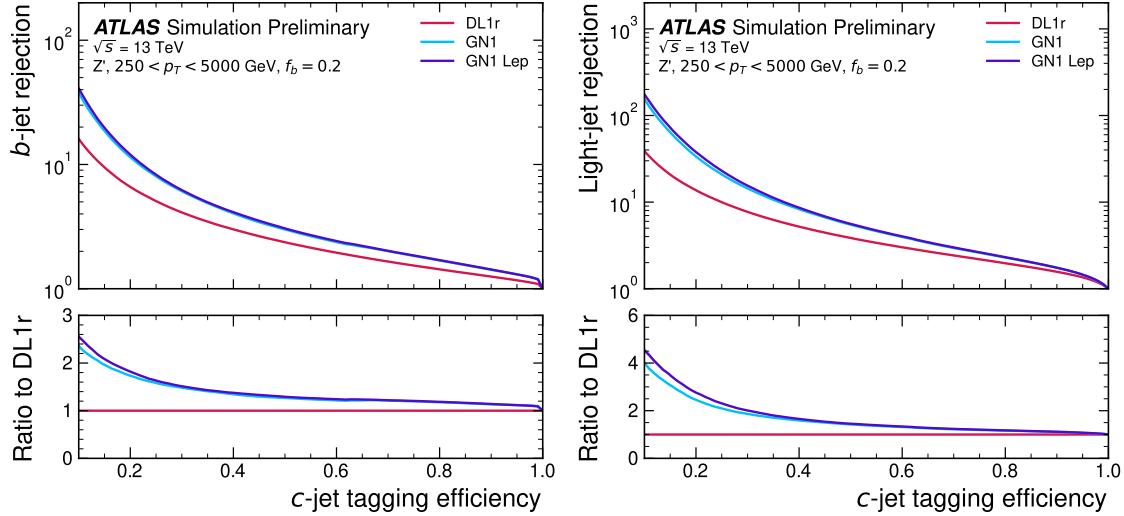
A value of  $f_b = 0.2$  is used for all models. Similar to Section 5.5.1, performance of the different taggers is compared by scanning through a range of  $c$ -jet tagging efficiencies and plotting the corresponding  $b$ - and light-jet rejection rates. As in Section 5.5.1, WPs are defined using jets 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, improving the  $b$ -rejection by 60% and the light-jet rejection by a factor of 2 at the 25%  $c$ -jet WP.

### 5.5.3 Ablations

Several ablations, the removal of components in the model to study their impact, are carried out to determine the importance of the auxiliary training objectives of GN1 to the overall performance. The “GN1 No Aux” variant retains the primary jet classification objective, but removes both track classification and vertexing auxiliary objectives (see Section 5.4.2) and as such only minimises the jet classification loss.



**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.

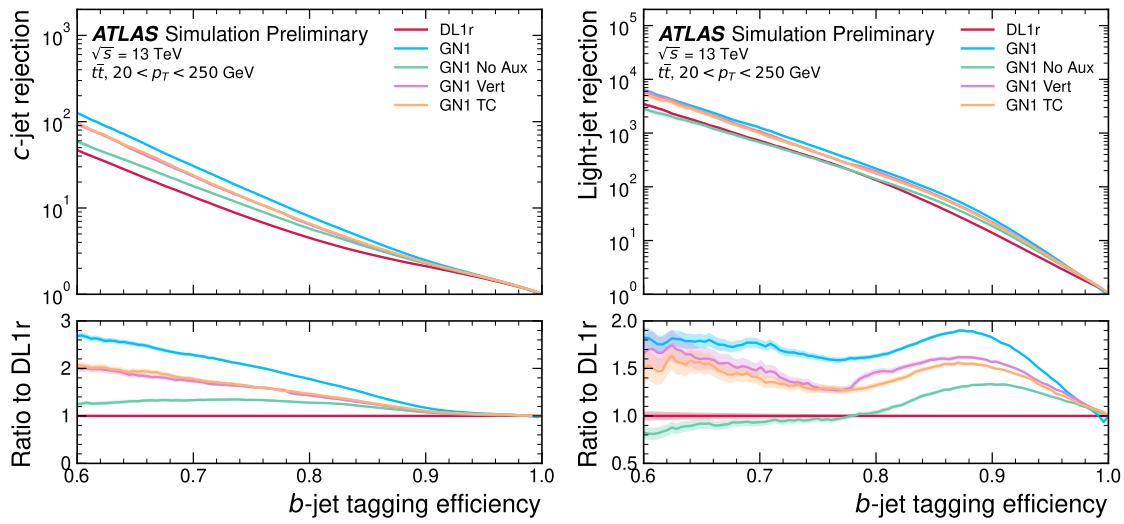
1352 The “GN1 TC” variant includes track classification but not vertexing, while “GN1  
1353 Vert” includes vertexing, but not track classification.

1354 For jets in both the  $t\bar{t}$  and  $Z'$  samples, the models without one or both of the auxiliary  
1355 objectives display significantly reduced  $c$ - and light-jet rejection when compared with  
1356 the baseline GN1 model, as shown in Figs. 5.10 and 5.11. For jets in the  $t\bar{t}$  sample,  
1357 the performance of GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert  
1358 perform similarly to each other. For jets in the  $Z'$  sample, the GN1 No Aux model  
1359 shows a clear improvement in  $c$ - and light-jet rejection when compared with DL1r at  
1360 lower  $b$ -jet tagging efficiencies. Similar to jets in the  $t\bar{t}$  sample, GN1 TC and GN1  
1361 Vert perform similarly, and bring large gains in background rejection when compared  
1362 with GN1 No Aux, but the combination of both auxiliary objectives yields the best  
1363 performance.

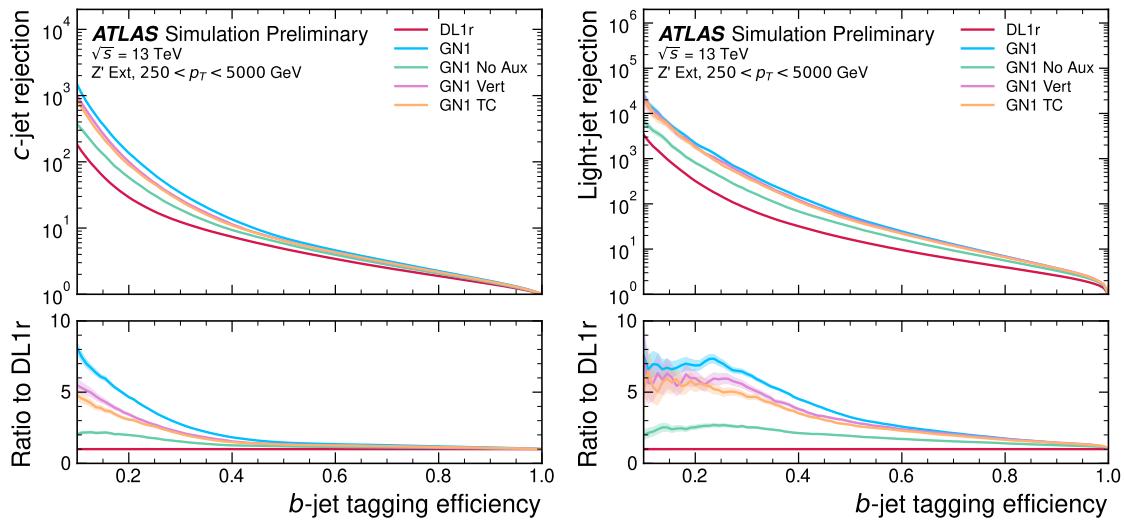
1364 It is notable that the GN1 No Aux model matches or exceeds the performance of  
1365 DL1r without the need for inputs from the low-level algorithms. This indicates that  
1366 the performance improvements enabled by GN1 appear to be able to compensate for  
1367 the removal of the low-level algorithm inputs. The GN1 TC and GN1 Vert variants  
1368 each similarly outperform DL1r, demonstrating that both contribute to the overall  
1369 high performance of the baseline model.

#### 1370 5.5.4 Inclusion of Low-Level Vertexing Algorithms

1371 GN1 does not include inputs from low-level tagging algorithms, including the vertexing  
1372 tools SV1 and JetFitter [45]. Since these algorithms are known to improve the  
1373 performance of DL1r, it was feasible that their inclusion in GN1 may further improve  
1374 on the performance of the GN1 models. In a dedicated training of GN1 the SV1 and  
1375 JetFitter tagger outputs were added to the GN1 jet classification network as an input,  
1376 similar to their use in DL1r. These outputs include information on the reconstructed  
1377 vertices, including the number of vertices, the vertex mass, displacement, and other  
1378 properties. In addition, the index of the reconstructed SV1 or JetFitter vertices were  
1379 included as two track-level inputs to GN1. The jet classification performance of this  
1380 GN1 model was not significantly different to the baseline model, and in some cases  
1381 the performance was slightly reduced. A dedicated look at the vertexing performance  
1382 of GN1 with some comparisons to SV1 and JetFitter is found in Section 5.5.5



**Figure 5.10:** 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 5.11:** 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.

### 5.5.5 Vertexing Performance

From the track-pair vertex prediction described in Section 5.4.2, tracks can be partitioned into compatible groups representing vertices (see [72]). As such, GN1 is able to be used to perform vertex “finding”, but not vertex “fitting”, i.e. the reconstruction of a vertex’s properties, which currently still requires the use of a dedicated vertex fitter. In order to study the performance of the different vertexing tools inside  $b$ -jets, the truth vertex label of the tracks, discussed in Section 5.4.2, are used. To estimate the efficiency with which GN1 manages to find vertices inclusively, vertices from GN1 containing tracks identified as coming from a  $b$ -hadron are merged together and compared to the inclusive truth decay vertices that result from a  $b$ -hadron decay (where if there are multiple distinct truth vertices from a  $b$ -hadron decay they are also merged together). Vertices are compared with the target truth vertex and the number of correctly and incorrectly assigned tracks is computed. Since secondary vertex information is only recovered for reconstructed tracks, an efficiency of 100% here denotes that all possible secondary vertices are recovered given the limited track reconstruction efficiency. A vertex is considered matched if it contains at least 65% of the tracks in the corresponding truth vertex, and has a purity of at least 50%. GN1 manages to achieve an inclusive reconstruction efficiency in  $b$ -jets of  $\sim 80\%$ , demonstrating that it effectively manages to identify the displaced vertices from  $b$ -hadron decays.

#### More detail

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

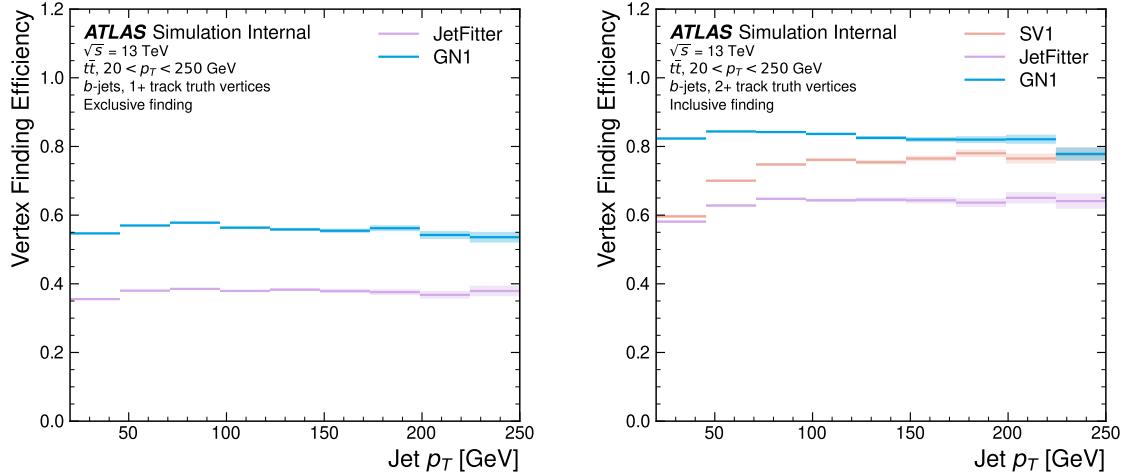
There are several caveats to a comparison of the vertexing tools which are a result of the different approaches they take to vertexing. SV1 and JetFitter are designed to only find secondary vertices in the jet, whereas GN1 is also trained to determine

1414 which tracks in the jet belong to the primary vertex (the vertex of the hard scatter  
 1415  $pp$  interaction). To account for this the GN1 vertex with the largest number of  
 1416 predicted primary tracks is excluded from the vertex finding efficiency calculation.  
 1417 While JetFitter and GN1 aim to resolve each displaced vertex inside the jet, such  
 1418 that secondary vertices from  $b$ -hadron decays are found separately to tertiary vertices  
 1419 from  $b \rightarrow c$  decay chains, SV1 by design attempts to find a single inclusive vertex  
 1420 per jet. This inclusive vertex groups inclusive  $b$ -hadron decays. These are tracks  
 1421 from the  $b$ -hadron decay itself (FromB) and tracks from  $b \rightarrow c$  decays (FromBC).  
 1422 In order to fairly compare the performance if the different tools, both the exclusive  
 1423 and inclusive vertex finding efficiency is studied. For the exclusive vertex finding  
 1424 case JetFitter and GN1 can be directly compared, while a comparison with SV1 is  
 1425 not possible due to aforementioned design constraints. The inclusive vertex finding  
 1426 performance of all three tools can be compared using the procedure outlined below.

1427 The starting point for the secondary vertex finding efficiency in both the exclusive  
 1428 and inclusive cases is to select truth secondary vertices are those containing only  
 1429 inclusive  $b$ -hadron decays to be considered as initial targets. For exclusive vertex  
 1430 finding, these truth secondary vertices can be used directly as the denominator for the  
 1431 efficiency calculation. Meanwhile for the inclusive efficiency all such truth secondary  
 1432 vertices in the jet are merged into a single inclusive target vertex. Correspondingly,  
 1433 for the inclusive vertex finding case, the vertices found by JetFitter are merged into  
 1434 a single vertex, and the vertices found by GN1 with at least one predicted inclusive  
 1435  $b$ -hadron decay track are also merged similarly. SV1 does not require any vertex  
 1436 merging.

1437 Next, in both cases for each truth secondary vertex, vertices in the jet found by the  
 1438 different vertexing tools are compared with the target truth vertex. The number  
 1439 of correctly and incorrectly assigned tracks is computed. In order to call a vertex  
 1440 efficient, it is required to contain at least 65% of the tracks in the corresponding  
 1441 truth vertex, and to have a purity of at least 50%. Single track vertices are required  
 1442 to have a purity of 100%. Additionally, for GN1 only, at least one track in the vertex  
 1443 is required to have a predicted heavy flavour origin.

1444 Vertex finding efficiencies for  $b$ -jets in the  $t\bar{t}$  sample are displayed as a function of  $p_T$   
 1445 separately for the inclusive and exclusive approaches in Fig. 5.12. For  $b$ -jets in the  $t\bar{t}$   
 1446 sample with  $20 < p_T < 250$  GeV, the exclusive vertex finding efficiency of JetFitter  
 1447 and GN1 is relatively flat as a function of  $p_T$ . Of the truth secondary vertices in this

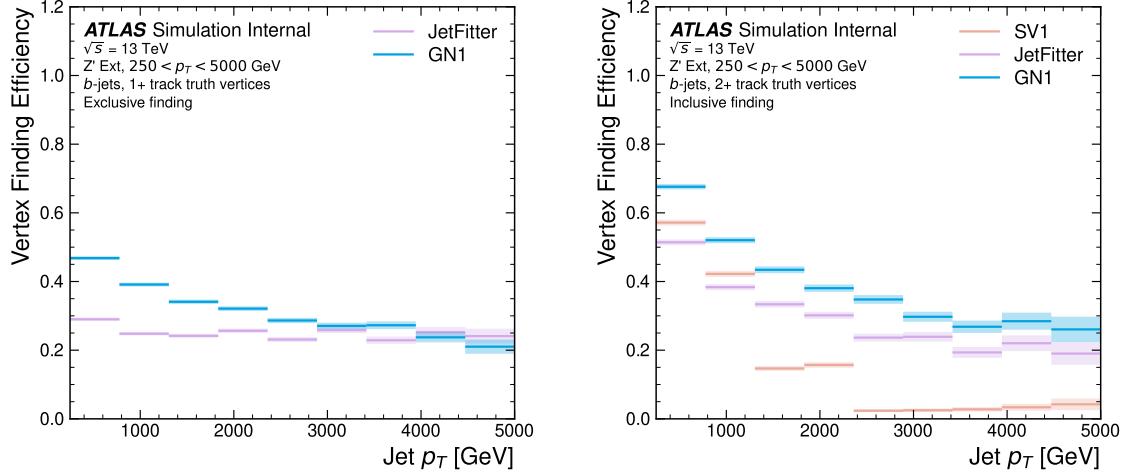


**Figure 5.12:** Vertex finding efficiency as a function of jet  $p_T$  for  $b$ -jets in the  $t\bar{t}$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.

1448  $p_T$  region, JetFitter efficiently finds approximately 40% and GN1 finds approximately  
 1449 55%. When finding vertices inclusively the vertex finding efficiency is generally higher.  
 1450 An increased dependence on  $p_T$  is also visible for JetFitter and SV1. As the jet  $p_T$   
 1451 increases from 20 GeV to 100 GeV, the efficiency of JetFitter increases from 55% to  
 1452 65%. In the same range, the efficiency of SV1 increases from 55% to 75%. GN1  
 1453 displays less dependence on  $p_T$  than JetFitter and SV1, efficiently finding upwards  
 1454 of 80% of vertices in  $b$ -jets in this  $p_T$  region. For  $b$ -jets with  $p_T > 100$  GeV, JetFitter  
 1455 finds approximately 65% of vertices, SV1 finds approximately 75% of vertices, and  
 1456 GN1 finds approximately 80% of vertices.

1457 For  $b$ -jets in the  $Z'$  sample, the vertex finding efficiency drops steeply with increasing  
 1458  $p_T$  up until  $p_T = 3$  TeV. GN1 outperforms SV1 and JetFitter across the  $p_T$  spectrum.  
 1459 In the first bin, the efficiency of GN1 is 75%, while the efficiencies of SV1 and JetFitter  
 1460 are around 60%. The efficiency of SV1 drops rapidly to almost zero above 3 TeV,  
 1461 while JetFitter and GN1 retain approximately 30% efficiency. Fig. 5.13 compares  
 1462 the exclusive vertex finding efficiencies of JetFitter and GN1 for multitrack vertices.  
 1463 JetFitter finds 45-50% of vertices in  $b$ -jets in the  $t\bar{t}$  sample, while GN1 finds 60-65%.  
 1464 For  $b$ -jets in the  $Z'$  sample, JetFitter finds 35% of vertices in the first bin, dropping

<sup>1465</sup> to 20% of vertices above 2 TeV. GN1 finds 55% of vertices in the first bin, dropping  
<sup>1466</sup> to 30% above 2 TeV.



**Figure 5.13:** 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.

### <sup>1467</sup> 5.5.6 Track Classification Performance

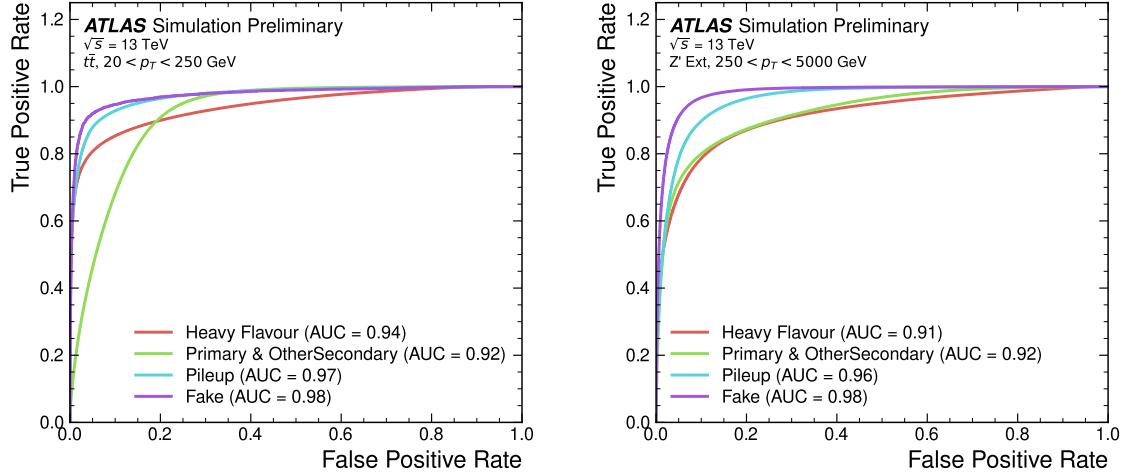
<sup>1468</sup> As discussed in Section 5.4.2, one of the auxiliary training objectives for GN1 is to  
<sup>1469</sup> predict the truth origin of each track in the jet. Since the equivalent information is  
<sup>1470</sup> not provided by any of the existing flavour tagging tools, as a benchmark a multi-class  
<sup>1471</sup> classification multilayer perceptron (MLP) is trained on the same tracks used for  
<sup>1472</sup> the baseline GN1 training. The model uses the same concatenated track-and-jet  
<sup>1473</sup> inputs as GN1 (see Section 5.4.1), but processes only a single track at a time. The  
<sup>1474</sup> model is comprised of five densely connected layers with 200 neurons per layer,  
<sup>1475</sup> though the performance was not found to be strongly sensitive to changes in the  
<sup>1476</sup> network structure. To measure the track classification performance, the area under  
<sup>1477</sup> the curve (AUC) of the receiver operating characteristic (ROC) curve is computed  
<sup>1478</sup> for each origin class using a one versus all classification approach. The AUCs for the  
<sup>1479</sup> different truth origin classes are averaged using both an unweighted and a weighted  
<sup>1480</sup> approach. The unweighted mean treats the performance of each class equally, while  
<sup>1481</sup> the weighted mean uses the fraction of tracks from each origin as a weight. As seen

in Table 5.4, 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.

**Table 5.4:** 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.

		AUC	
		Mean	Weighted
$t\bar{t}$	MLP	0.87	0.89
	GN1	<b>0.92</b>	<b>0.95</b>
$Z'$	MLP	0.90	0.94
	GN1	<b>0.94</b>	<b>0.96</b>

Fig. 5.14 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.



**Figure 5.14:** 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.

## 1499 5.6 Conclusion

1500 A novel jet tagger, GN1, with a graph neural network architecture and trained  
 1501 with auxiliary training targets, is presented and now fully implemented in the  
 1502 ATLAS software. GN1 is shown to improve flavour tagging performance with respect  
 1503 to DL1r, the current default ATLAS flavour tagging algorithm, when compared  
 1504 in simulated collisions. GN1 improves  $c$ - and light-jet rejection for jets in the  $t\bar{t}$   
 1505 sample with  $20 < p_T < 250 \text{ GeV}$  by factors of  $\sim 2.1$  and  $\sim 1.8$  respectively at a  $b$ -jet  
 1506 tagging efficiency of 70% when compared with DL1r. For jets in the  $Z'$  sample  
 1507 with  $250 < p_T < 5000 \text{ GeV}$ , GN1 improves the  $c$ -rejection by a factor of  $\sim 2.8$   
 1508 and light-jet rejection by a factor of  $\sim 6$  for a comparative  $b$ -jet efficiency of 30%.  
 1509 Previous multivariate flavour tagging algorithms relied on inputs from low-level  
 1510 tagging algorithms, whereas GN1 needs no such inputs, making it more flexible.  
 1511 It can be easily fully optimised via a retraining for specific flavour tagging use  
 1512 cases, as demonstrated with  $c$ -tagging and high- $p_T$   $b$ -tagging, without the need for  
 1513 time-consuming retuning of the low-level tagging algorithms. The model is also  
 1514 simpler to maintain and study due to the reduction of constituent components. GN1  
 1515 demonstrates improved track classification performance when compared with a simple

1516 per-track MLP and an efficiency of  $\sim 80\%$  for inclusive vertex finding in  $b$ -jets. The  
1517 auxiliary track classification and vertex finding objectives are shown to significantly  
1518 contribute to the performance in the jet classification objective, and are directly  
1519 responsible for the improvement over DL1r. Further studies need to be undertaken  
1520 to verify the performance of GN1 on collision data.

1521 **5.7 Extensions**

1522 **5.7.1 Looser Track Selection**

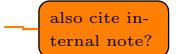
<sub>1523</sub>

# Chapter 6

<sub>1524</sub>

## Boosted VHbb Analysis

<sub>1525</sub> This work has been published in Ref. [96]. Figures and tables are reproduced here.



<sub>1526</sub>

### 6.1 Overview

<sub>1527</sub> Lifted from old Vhbb preamble chapter:

<sub>1528</sub> The Higgs boson, discovered at the LHC in 2012, is predicted by the standard model  
<sub>1529</sub> to decay primarily to two  $b$  quarks, with a branching factor of  $0.582 \pm 0.007$  [24].  
<sub>1530</sub> Observation of this decay mode was recently reported by ATLAS [66]. Whilst the  
<sub>1531</sub> dominant Higgs production mode at the LHC is gluon-gluon fusion, this mode has  
<sub>1532</sub> an overwhelming QCD multijet background and so sensitivity to the Higgs is low.  
<sub>1533</sub> The  $H \rightarrow b\bar{b}$  observation therefore searched for Higgs bosons produced in association  
<sub>1534</sub> with a vector boson (W or Z). This production mechanism results in leptonic final  
<sub>1535</sub> states from the decay of the vector boson, allowing for leptonic triggering, whilst at  
<sub>1536</sub> the same time significantly reducing the multi-jet background.

<sub>1537</sub> A closely related analyses now searches for the  $H \rightarrow b\bar{b}$  decay of the Higgs boson,  
<sub>1538</sub> produced in association with a vector boson, when the vector boson and Higgs are  
<sub>1539</sub> highly boosted. The full Run-2 dataset is used for a total integrated luminosity of  
<sub>1540</sub>  $139 \text{ fb}^{-1}$ . The analysis is split into 0-, 1- and 2-lepton channels depending on the  
<sub>1541</sub> number of selected electrons and muons, to target the  $ZH \rightarrow \nu\nu bb$ ,  $WH \rightarrow \ell\nu bb$ ,  
<sub>1542</sub>  $ZH \rightarrow \ell\ell bb$  processes, respectively, where  $\ell$  is an electron or muon. In all channels,  
<sub>1543</sub> events are required to have exactly two  $b$ -tagged jets, which form the Higgs boson  
<sub>1544</sub> candidate. At least one of the  $b$ -tagged jets is required to have  $p_T$  greater than 45

1545 GeV. Events are further split into 2-jet or 3-jet categories depending on whether  
1546 additional, untagged jets are present.

1547 In the 0- and 1-lepton channels, the analysis is further split into signal and control  
1548 regions. To leading order, there are no additional  $b$ -jets in the event other than  
1549 the two coming from the reconstructed Higgs candidate. For this reason, there is  
1550 a signal region veto (i.e. events are not accepted into the signal region) for events  
1551 with additional  $b$ -tagged jets in the event. Events with additional  $b$ -tagged jets are  
1552 included in the control region, which is highly pure in  $t\bar{t}$  events. The control region  
1553 is used to constrain the normalisation of the  $t\bar{t}$  background.

## 1554 6.2 Introduction

1555 Lifted from paper:

1556 Since the discovery of the Higgs boson ( $H$ ) [9–11, 97] with a mass of around 125 GeV  
1557 [98] by the ATLAS and CMS Collaborations [12, 13] in 2012, the analysis of proton–  
1558 proton ( $pp$ ) collision data at centre-of-mass energies of 7 TeV, 8 TeV and 13 TeV  
1559 delivered by the Large Hadron Collider (LHC) [65] has led to precise measurements  
1560 of the main production cross-sections and decay rates of the Higgs boson, as well  
1561 as measurements of its mass and its spin and parity properties. In particular, the  
1562 observation of the decay of the Higgs boson into  $b$ -quark pairs provided direct  
1563 evidence for the Yukawa coupling of the Higgs boson to down-type quarks [26, 66].  
1564 Finally, a combination of 13 TeV results searching for the Higgs boson produced in  
1565 association with a leptonically decaying  $W$  or  $Z$  boson established the observation  
1566 of this production process [66]. A first cross-section measurement as a function  
1567 of the vector-boson transverse momentum was also carried out by the ATLAS  
1568 Collaboration [99].

1569 The previous ATLAS analyses [66, 99] in this channel were mainly sensitive to vector  
1570 bosons with transverse momentum ( $p_T$ ) in the range of approximately 100–300 GeV.  
1571 These analyses considered a pair of jets with radius parameter of  $R = 0.4$ , referred  
1572 to as small-radius (small- $R$ ) jets, to reconstruct the Higgs boson. For higher Higgs  
1573 boson transverse momenta, the decay products can become close enough that they  
1574 cannot be reconstructed with two small- $R$  jets. To explore this ‘boosted’ regime,

the Higgs boson is reconstructed as a single large- $R$  jet with  $R = 1.0$  [100]. This high- $p_{\mathrm{T}}$  regime is particularly interesting due to its sensitivity to physics beyond the Standard Model [101].

This Letter presents a measurement of cross-sections for the associated production of a high transverse momentum Higgs boson that decays into a  $b\bar{b}$  pair with a leptonically decaying  $W$  or  $Z$  boson. The analysis uses  $pp$  collision data recorded between 2015 and 2018 by the ATLAS detector [30] during Run 2 at the LHC. This dataset corresponds to an integrated luminosity of  $139 \text{ fb}^{-1}$ . Events are selected in 0-, 1- and 2-lepton channels, based on the number of reconstructed charged leptons,  $\ell$  (electrons or muons), in the final state to explore the  $ZH \rightarrow \nu\nu b\bar{b}$ ,  $WH \rightarrow \ell\nu b\bar{b}$  and  $ZH \rightarrow \ell\ell b\bar{b}$  signatures, respectively. The Higgs boson is reconstructed as a single large- $R$  jet and the  $b$ -quarks from its decay as a pair of jets, reconstructed with a  $p_{\mathrm{T}}$ -dependent radius parameter, associated with the large- $R$  jet and identified as containing a  $b$ -hadron.

The analysis using small- $R$  jets and focusing on slightly lower Higgs boson transverse momentum regions was recently updated with the complete Run 2 dataset [102]. The large- $R$  jet analysis significantly overlaps with the small- $R$  jets analysis. The two results can therefore not be straightforwardly combined.

The dominant background processes after the event selection correspond to the production of  $V + \text{jets}$ , where  $V$  refers to either a  $W$  or  $Z$  boson,  $t\bar{t}$ , single-top and dibosons. The signal is extracted from a combined profile likelihood fit to the large- $R$  jet mass, using several signal and control regions. The yield of diboson production  $VZ$  with  $Z \rightarrow b\bar{b}$  is also measured using the same fit and provides a validation of the analysis. The cross-section measurements are performed within the simplified template cross-section (STXS) framework [24, 103]. These measurements are then used to constrain anomalous couplings in a Standard Model effective field theory (SMEFT) [104].

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

**Table 6.1:** 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.

## 1602 6.3 Modelling Work

### 1603 6.3.1 Background

#### 1604 Alternative Samples

1605 As mentioned, alternative samples of V+jets events was generated using MAD-  
 1606 GRAPH5\_AMC@NLO+PYTHIA8, and the results are compared with the nominal  
 1607 SHERPA 2.2.1 samples. This allows for a comparison of different parton showering  
 1608 and underlying event models, and derivation of the systematic uncertainties on the  
 1609 nominal choice of models.

#### 1610 Internal Weight Variations

1611 Nominal signal samples generated with SHERPA 2.2.1 include systematic variations  
 1612 of certain modelling parameters which are stored as alternative event weights. The  
 1613 samples contain event weight variations which correspond to variations of renormal-  
 1614 isation scale  $\mu_R$ , and factorisation scale  $\mu_F$ , of 0.5 and 2 times the nominal value.  
 1615 Additionally stored is event weight variations corresponding to 30 different variations  
 1616 on the PDF and two variations of the strong coupling constant  $\alpha_S$ . Variations of

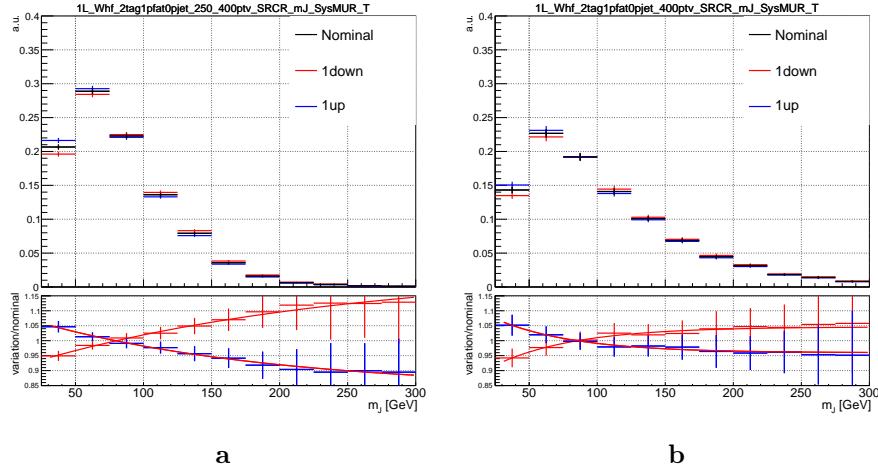
1617  $\alpha_S$  were found to have negligible impact on the results of the analysis, and are not  
1618 discussed further.

1619 **Parameterisation Methods**

1620 While the inclusion of internal weight variation in MC event generators has decreased  
1621 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently  
1622 some sources of systematic uncertainty that are unable to be stored as internal weight  
1623 variations due to technical limitations. Two such systematics relate to the choice of  
1624 CKKW matrix element merging scale, and resummation scale (QSF). The generation  
1625 of high statistics alternative samples is a time consuming process, as is typically  
1626 not done for all samples for every new generator release. A method to parameterise  
1627 the systematic variation using one sample, and to then apply this parameterisation  
1628 to another sample, has been developed by the ATLAS SUSY group [105]. This  
1629 method was used to derive CKKW and QSF uncertainties for the nominal SHERPA  
1630 2.2.1 sample, using a previous (lower statistic) SHERPA 2.1 alternative sample. The  
1631 resulting uncertainties were studied and found to be negligible in comparison with  
1632 systemics from other sources.

1633 **Shape Uncertainties**

1634 In order to derive shape uncertainties (which as the name suggests affect shapes  
1635 but not overall normalisations of distributions), the following procedure is carried  
1636 out. Normalised distributions of the reconstructed Higgs candidate mass  $m_J$  are  
1637 compared for the nominal sample and variations. For each variation, the ratio of the  
1638 variation to nominal is calculated, and an analytic function is fit to those sources  
1639 of variation which have a ratio deviating from unity. If different analysis regions or  
1640 channels show the same pattern of variation, a common uncertainty is assigned. An  
1641 example of a significant source of uncertainty, arising from choice of factorisation  
1642 scale  $\mu_R$  is shown in Fig. 6.1. An exponential function has been fitted to the ratio  
1643 of the normalised distributions. Two different analysis regions (medium and high  
1644  $p_{T^V}$  bins) are shown. The difference of the shape of the variation means that two  
1645 separate uncertainties have to be added in the fit, and applied individually in each  
1646  $p_{T^V}$  region.



**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  $\mu_R$  has been varied by a factor of 2 (“1up”) and 0.5 (“1down”). An exponential function has been fit to the ratio.

### 1647 Acceptance Uncertainties

1648 Several different types of acceptance uncertainties have been calculated. These  
 1649 are implemented as nuisance parameters in the fit and for the most part account  
 1650 for the migration of events between different analysis regions. The list acceptance  
 1651 uncertainties relevant to the V+jets processes are given summarised below.

- 1652 • **Overall normalisation:** only relevant where normalisation cannot be left  
 1653 floating (i.e. determined in the fit).
- 1654 • **SR-to-CR relative acceptance:** the uncertainty on the normalisation of the  
 1655 signal region due to events migrating between the signal and control regions.
- 1656 • **HP-to-LP relative acceptance:** the uncertainty on the normalisation of the  
 1657 high-purity (HP) signal region due to events migrating between the high- and  
 1658 low-purity signal regions.
- 1659 • **Medium-to-high  $p_{T^V}$  relative acceptance:** describes any ‘shape’ effect in  
 1660  $p_{T^V}$  distribution, given that the analysis only uses two  $p_{T^V}$  bins (medium and  
 1661 high).

- 1662 • **Flavour relative acceptance:** for each flavour  $V_{xx}$ , where  $xx \in \{bc, bl, cc\}$   
1663 the ratio of  $V_{xx}/V_{bb}$  events is calculated. This corresponds to the uncertainty  
1664 of  $V_{bb}$  events due to the miss-tagging of other flavours  $V_{xx}$ .

1665 The uncertainties on different systematics are summed in quadrature to give a total  
1666 uncertainty on each region. A summary of the different acceptance uncertainties that  
1667 were derived in this way and subsequently applied in the fit are given in Table 6.2.  
1668 An effort has been made, wherever possible, to harmonise similar uncertainties across  
1669 different analysis regions and channels.

### 1670 6.3.2 Vector Boson + Jets Modelling

1671 The background processes involving  $W$  or  $Z$  boson decays into leptons (including  
1672 those in which the  $W$  boson arises from a top-quark decay) are collectively referred  
1673 to as electroweak (EW), or  $V+jets$ , backgrounds.  $W+jets$  events are most relevant  
1674 to the 1-lepton channel via the leptonic decay of  $W \rightarrow \ell\nu$ . In the event of  $W \rightarrow \tau\nu$ ,  
1675 and subsequent decay of the  $\tau$ , or the lack of the successful reconstruction of the  
1676  $e$  or  $\mu$ ,  $W+jets$  can also contribute to the 0-lepton channel. Meanwhile,  $Z+jets$   
1677 contributes primarily to the 0- and 2-lepton channels via the processes  $Z \rightarrow \nu\nu$  and  
1678  $Z \rightarrow \ell\ell$  respectively.

1679 Modelling is used to predict the outcomes of the analysis and to assess the impact of  
1680 sources of different systematic uncertainty. Signal and background modelling has  
1681 has primarily consisted of using Monte Carlo (MC) generators to produce simulated  
1682 events. The uncertainties on the simulated output must be well understood to  
1683 perform a successful analysis. To achieve this, a set of “nominal” samples are first  
1684 defined as a reference to which different variations can be compared. The nominal  
1685 samples are chosen as the best possible representation of the underlying physical  
1686 process. “Alternative” samples are used to understand the systematic uncertainties  
1687 on the nominal samples. To generate an alternative sample, some aspect of the model  
1688 is varied, and the simulation is re-run. A comparison back to the nominal sample  
1689 gives a handle on the systematic uncertainty associated with the model parameter  
1690 which was changed. Detailed information can be found in [106]. In order to access  
1691 uncertainties associated with the use of MC generators, variations of the data are  
1692 produced using alternative generators or variation of nominal generator parameters.  
1693 The variation of nominal generator parameters can in certain cases be implemented

1694 using internal weight variations stored alongside the nominal events, and in other  
 1695 cases a new independent sample must be generated. The nominal generator used  
 1696 for V+jets events is SHERPA 2.2.1, while MADGRAPH5\_AMC@NLO+PYTHIA8  
 1697 (which uses different parton showering models) is used as an alternative generator.  
 1698 As production of large MC samples is computationally expensive, a feature of state  
 1699 of the art simulation packages is to store some sources of variation as internal event  
 1700 weights, which can be generated alongside the nominal samples, saving computation  
 1701 time. Several sources of uncertainty, summarised in Table 6.1, have been assessed.

V+jets Acceptance Uncertainties				
Boson	W		Z	
Channel	0L	1L	0L	2L
Vbb Norm.	30%	-	-	-
SR/CR	90% <sup>†</sup>	40% <sup>†</sup>	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 6.2:** 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 Zbb norm is applied in the 1L channel when a floating normalisation for Zbb cannot be obtained from the 0L or 2L channels.

<sup>1702</sup> **6.3.3 Diboson Modelling**

<sup>1703</sup> **6.4 Fit Studies**

<sup>1704</sup> **6.4.1 Fit Model**

<sup>1705</sup> A global profile likelihood fit is used to extract the signal strength  $\mu$  and its significance  
<sup>1706</sup> from the data. This statistical setup treats each bin as a Poisson counting experiment.  
<sup>1707</sup> The combined likelihood over  $N$  bins, without considering sources of systematic  
<sup>1708</sup> uncertainty, is given by

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

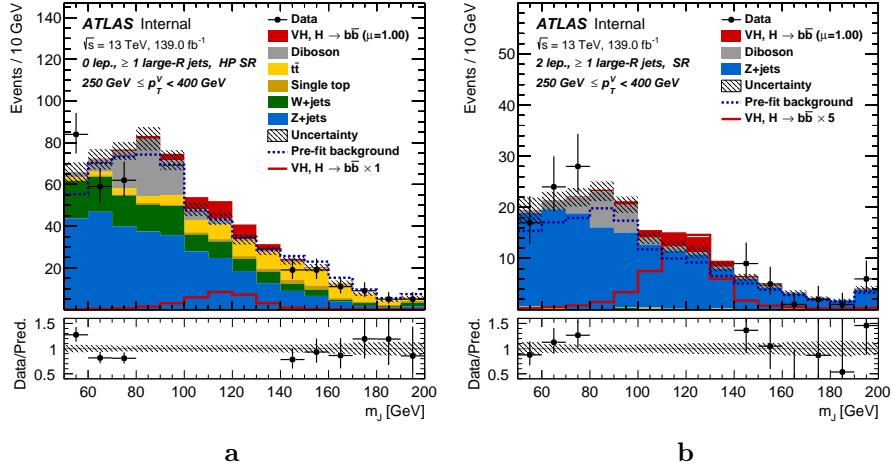
<sup>1709</sup> where  $s_i$  ( $b_i$ ) is the expected number of signal (background) events in bin  $i$ , and  
<sup>1710</sup>  $n_i$  is the number of events observed in data in bin  $i$ . The presence of systematic  
<sup>1711</sup> uncertainties which can affect the expected numbers of signal and background events  
<sup>1712</sup> necessitates the addition of nuisance parameters (NPs),  $\theta$ , to the likelihood. Each  
<sup>1713</sup> source of systematic uncertainty for V+jets samples discussed in the previous section  
<sup>1714</sup> was implemented as a NP  $\theta_j$  in the fit. The presence of NPs modifies the likelihood  
<sup>1715</sup> as

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

<sup>1716</sup> where

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

<sup>1717</sup> Post-fit  $m_J$  distributions in the high-purity medium  $p_T V$  regions for the 0- and  
<sup>1718</sup> 2-lepton channels are shown in Fig. 6.2. The plots show large falling backgrounds,  
<sup>1719</sup> predominantly made up of  $W$ +jets and  $Z$ +jets events, and a signal distribution  
<sup>1720</sup> corresponding to the Standard Model Higgs boson peaking around  $m_H = 125$  GeV.



**Figure 6.2:** Post-fit distributions for the 0-lepton (6.2a) and 2-lepton (6.2b) 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.

## 6.5 Conclusion

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

<sub>1731</sub> Chapter 7

<sub>1732</sub> Conclusion

<sup>1733</sup> Appendix A

<sup>1734</sup> Combining Multiple Triggers

<sub>1735</sub>

# Bibliography

- <sub>1736</sub> [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- <sub>1737</sub>
- <sub>1738</sub> [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.
- <sub>1739</sub>
- <sub>1740</sub>
- <sub>1741</sub> [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.
- <sub>1742</sub>
- <sub>1743</sub>
- <sub>1744</sub> [4] CDF collaboration, *Observation of top quark production in  $\bar{p}p$  collisions*, *Phys. Rev. Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- <sub>1745</sub>
- <sub>1746</sub> [5] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) pp. 2632–2637 [[hep-ex/9503003](#)].
- <sub>1747</sub>
- <sub>1748</sub> [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.
- <sub>1749</sub>
- <sub>1750</sub> [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.
- <sub>1751</sub>
- <sub>1752</sub>
- <sub>1753</sub> [8] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) pp. 218–224 [[hep-ex/0012035](#)].
- <sub>1754</sub>
- <sub>1755</sub> [9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- <sub>1756</sub>
- <sub>1757</sub> [10] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) pp. 508–509.
- <sub>1758</sub>
- <sub>1759</sub> [11] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*

- 1760        *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 1761        [12] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*  
1762        *Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)  
1763        p. 1 [[1207.7214](#)].
- 1764        [13] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*  
1765        *experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 1766        [14] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)  
1767        p. 083C01.
- 1768        [15] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge*  
1769        *invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 1770        [16] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)  
1771        pp. 579–588.
- 1772        [17] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 1773        [18] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel*  
1774        *symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**  
1775        (1968) pp. 367–377.
- 1776        [19] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,  
1777        *Phys. Rev.* **104** (1956) pp. 254–258.
- 1778        [20] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental*  
1779        *test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 1780        [21] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of*  
1781        *conservation of parity and charge conjugation in meson decays: the magnetic moment*  
1782        *of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 1783        [22] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,  
1784        *Phys. Rev.* **117** (1960) pp. 648–663.
- 1785        [23] J. Goldstone, *Field theories with «superconductor» solutions*, *Il Nuovo Cimento*  
1786        (1955–1965) **19** (1961) pp. 154–164.
- 1787        [24] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs*  
1788        *Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](#).
- 1789        [25] ATLAS Collaboration, *Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the*

- 1790        *ATLAS detector*, ATLAS-CONF-2018-036 (2018).
- 1791 [26] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys. Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](#)].
- 1793 [27] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) pp. S08002–S08002.
- 1795 [28] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) p. S08004.
- 1797 [29] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of Instrumentation* **3** (2008) pp. S08005–S08005.
- 1799 [30] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) p. S08003.
- 1801 [31] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 1804 [32] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 1807 [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](#).
- 1810 [34] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, *JINST* **11** (2016) p. P04008 [[1512.01094](#)].
- 1812 [35] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 1814 [36] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 1816 [37] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 1818 [38] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) p. T05008 [[1803.00844](#)].

- 1820 [39] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys. J. C* **77** (2017) p. 317 [[1611.09661](#)].
- 1821
- 1822 [40] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, (2021).
- 1823 [41] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 1824
- 1825 [42] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense Environments*, ATL-PHYS-PUB-2015-006 (2015).
- 1826
- 1827 [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](#)].
- 1828
- 1829 [44] ATLAS collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, *arXiv e-prints* (2014) p. arXiv:1406.7690 [[1406.7690](#)].
- 1830
- 1831 [45] 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](#)].
- 1832
- 1833
- 1834 [46] M. Cacciari, G. P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) p. 063 [[0802.1189](#)].
- 1835
- 1836 [47] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 1837
- 1838 [48] 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](#)].
- 1839
- 1840
- 1841 [49] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018 (2014).
- 1842
- 1843 [50] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012).
- 1844
- 1845 [51] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024 (2016).
- 1846
- 1847 [52] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000) pp. 577–606 [[hep-ph/9912292](#)].
- 1848
- 1849 [53] A. Chen, M. Goldberg, N. Horwitz, A. Jawahery, P. Lipari, G. C. Moneti et al.,

- 1850        *Limit on the  $b \rightarrow u$  coupling from semileptonic  $b$  decay*, *Phys. Rev. Lett.* **52** (1984)  
1851        pp. 1084–1088.
- 1852        [54] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS*  
1853        *Experiment*, Master's thesis, Innsbruck U., 2013.
- 1854        [55] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*  
1855        *Process Interface for Event Generators*, *ArXiv High Energy Physics - Phenomenology*  
1856        *e-prints* (2001) .
- 1857        [56] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*  
1858        *standard format for Les Houches Event Files*, *Computer Physics Communications*  
1859        **176** (2007) pp. 300–304.
- 1860        [57] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*  
1861        *universal approximators*, *Neural Networks* **2** (1989) pp. 359–366.
- 1862        [58] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*  
1863        *back-propagating errors*, *nature* **323** (1986) pp. 533–536.
- 1864        [59] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*  
1865        *activity*, *The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.
- 1866        [60] J. Hopfield, *Neural networks and physical systems with emergent collective*  
1867        *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*  
1868        *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 1869        [61] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*  
1870        *networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 1871        [62] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*  
1872        *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 1873        [63] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter*  
1874        *Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014 (2020).
- 1875        [64] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS*  
1876        *Detector*, ATL-PHYS-PUB-2022-027 (2022).
- 1877        [65] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 1878        [66] ATLAS Collaboration, *Observation of  $H \rightarrow b\bar{b}$  decays and VH production with the*  
1879        *ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].

- 1880 [67] ATLAS Collaboration, *Observation of Higgs boson production in association with a*  
1881 *top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173  
1882 [[1806.00425](#)].
- 1883 [68] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*  
1884 *using  $139\text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector*, *JHEP* **03**  
1885 (2020) p. 145 [[1910.08447](#)].
- 1886 [69] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**  
1887 (2018) p. 030001.
- 1888 [70] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,  
1889 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,  
1890 *arXiv preprint arXiv:1806.01261* (2018) .
- 1891 [71] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,  
1892 *Secondary vertex finding in jets with neural networks*, *The European Physical Journal C* **81** (2021) .
- 1894 [72] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,  
1895 *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. arXiv:2002.08772  
1896 [[2002.08772](#)].
- 1897 [73] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*  
1898 *b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013 (2017).
- 1899 [74] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*  
1900 *ATLAS detector*, ATL-PHYS-PUB-2017-011 (2017).
- 1901 [75] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*  
1902 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003 (2017).
- 1903 [76] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,  
1904 *Journal of High Energy Physics* **2004** (2004) p. 040–040.
- 1905 [77] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*  
1906 *carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* **2007** (2007)  
1907 p. 126–126.
- 1908 [78] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*  
1909 *shower simulations: the powheg method*, *Journal of High Energy Physics* **2007** (2007)  
1910 p. 070–070.

- 1911 [79] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*  
1912 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*  
1913 *Physics* **2010** (2010) .
- 1914 [80] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015)  
1915 p. 040 [[1410.8849](#)].
- 1916 [81] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,  
1917 ATL-PHYS-PUB-2016-020 (2016).
- 1918 [82] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An*  
1919 *introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159  
1920 [[1410.3012](#)].
- 1921 [83] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,  
1922 ATL-PHYS-PUB-2014-021 (2014).
- 1923 [84] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013)  
1924 p. 244 [[1207.1303](#)].
- 1925 [85] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*  
1926 **462** (2001) p. 152.
- 1927 [86] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**  
1928 (2010) p. 823 [[1005.4568](#)].
- 1929 [87] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.*  
1930 *Instrum. Meth. A* **506** (2003) p. 250.
- 1931 [88] ATLAS Collaboration, *Muon reconstruction performance in early  $\sqrt{s} = 13$  TeV data*,  
1932 ATL-PHYS-PUB-2015-037 (2015).
- 1933 [89] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*  
1934 *experiment using the 2015 and 2016 LHC proton–proton collision data at*  
1935  *$\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 1936 [90] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*  
1937 *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, *arXiv e-prints* (2020)  
1938 p. arXiv:2007.08294 [[2007.08294](#)].
- 1939 [91] J. Shlomi, P. Battaglia and J.-R. Vlimant, *Graph neural networks in particle physics*,  
1940 *Machine Learning: Science and Technology* **2** (2021) p. 021001.
- 1941 [92] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,

- 1942 [arXiv e-prints (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 1943 [93] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola, *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 1945 [94] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 1946 [95] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).  
<https://github.com/onnx/onnx>.
- 1949 [96] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson decaying into b-quarks with a vector boson at high transverse momentum in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021) p. 136204 [[2008.02508](#)].
- 1953 [97] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) pp. 132–133.
- 1955 [98] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments*, *Phys. Rev. Lett.* **114** (2015) p. 191803 [[1503.07589](#)].
- 1958 [99] ATLAS Collaboration, *Measurement of VH,  $H \rightarrow b\bar{b}$  production as a function of the vector-boson transverse momentum in 13 TeV pp collisions with the ATLAS detector*, *JHEP* **05** (2019) p. 141 [[1903.04618](#)].
- 1961 [100] 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](#)].
- 1964 [101] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039 [[1512.02572](#)].
- 1967 [102] ATLAS Collaboration, *Measurements of WH and ZH production in the  $H \rightarrow b\bar{b}$  decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) p. 178 [[2007.02873](#)].
- 1970 [103] J. R. Andersen et al., *Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report*, in *9th Les Houches Workshop on Physics at TeV Colliders (PhysTeV 2015) Les Houches, France, June 1-19, 2015, 2016*, [1605.04692](#).

- 1973 [104] R. Contino, M. Ghezzi, C. Grojean, M. Mühlleitner and M. Spira, *Effective*  
1974       *Lagrangian for a light Higgs-like scalar*, *JHEP* **07** (2013) p. 035 [[1303.3876](#)].
- 1975 [105] J. K. Anders and M. D’Onofrio, *V+Jets theoretical uncertainties estimation via a*  
1976       *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 1977 [106] A. S. Bell and F. Lo Sterzo, *Signal and Background Modelling Studies for the*  
1978       *Standard Model VH,  $H \rightarrow b\bar{b}$  Analysis*, Tech. Rep. ATL-COM-PHYS-2018-505,  
1979       Geneva (2018).

# 1980 List of figures

<small>1981</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>1986</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>1990</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>1995</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. Reproduced from Ref. [33]. . . . .	24
<small>1998</small>	2.2 Delivered, recorded, and usable integrated luminosity as a function of time during Run 2 [32]. . . . .	26
<small>2000</small>	2.3 Average pile-up profiles measured by ATLAS during Run 2 [32]. During Run 3, even higher levels of pile-up are expected. . . . .	27
<small>2002</small>	2.4 A 3D model of the entire ATLAS detector. Cutouts through the detector .	27
<small>2003</small>	2.5 A 3D model of the ATLAS ID showing the barrel layers and end-cap disks.	28

2004	2.6 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown. . . . .	29
2005		
2006	2.7 A cross-sectional view of the ATLAS IBL. . . . .	30
2007		
2008	2.8 The ATLAS calorimeters. The ECal (orange) and HCal (grey, brown). . . . .	31
2009		
2010	2.9 The ATLAS muon spectrometer. . . . .	33
2011		
2012	2.10 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. . . . .	37
2013		
2014	2.11 A sketch of electron reconstruction using the ATLAS detector [51]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal. . . . .	39
2015		
2016	3.1 The truth $b$ -hadron decay radius $L_{xy}$ as a function of truth transverse momentum $p_T$ for reconstructed $b$ -jets in $Z'$ events. Error bars show the standard deviation. The pixel layers are shown in dashed horizontal lines. . . . .	43
2017		
2018		
2019	3.2 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. . . . .	43
2020		
2021		
2022		
2023		
2024	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. . . . .	44
2025		
2026		
2027		
2028		
2029		

2030	3.4 Hit multiplicities on the IBL (left) and the pixel layers (right) as a function 2031 of the $p_T$ of the reconstructed track. Tracks from the weak decay of the 2032 $b$ -hadron are shown in red, while fragmentation tracks (which are prompt) 2033 are in blue. For each of these, standard tracks and pseudo-tracks are plotted. 2034 Hit multiplicities on the pseudo-tracks at high $p_T$ due to the increased flight 2035 of the $b$ -hadron. The baseline tracks have more hits than the pseudo-tracks, 2036 indicating that they are being incorrectly assigned additional hits. . . . .	45
2037	3.5 $b$ -hadron decay track reconstruction efficiency. . . . .	45
2038	3.6 The rate of shared hits on $b$ -hadron decay tracks on the IBL (left), and 2039 the B-layer (right), as a function of the production radius of the $b$ -hadron 2040 decay product. Pseudo-tracks represent ideal performance given the ATLAS 2041 detector, and given also the efficiency of the NN cluster splitting algorithms.	48
2042	3.7 Profiles, as a function of parent $b$ -hadron $p_T$ , of good (left) and wrong (right) 2043 hit assignment rates on the IBL for tracks using baseline tracking (black), 2044 the modified version of the outlier removal procedure (red). . . . .	50
2045	3.8 (left) B hadron decay track $d_0$ pulls for baseline and modified GX2F tracks. 2046 (right) The magnitude of the decay track $d_0$ pull as a function of B hadron 2047 transverse momentum. . . . .	51
2048	4.1 A diagram displaying the logical flow of a single neuron with three inputs 2049 $x_i^j$ . Each input is multiplied by a weight $\theta_j$ , and the resulting values are 2050 summed. A bias term $\theta_0$ is added, and the result is passed to an activation 2051 function. Each neuron can be thought of as a logistic regression model. . .	54
2052	4.2 Several common choices for the activation function $g$ of an artificial neuron. .	55
2053	4.3 Rate of fake tracks as a function of jet transverse momentum (left) and 2054 $\Delta R(\text{track}, \text{jet})$ (right). The rate of fake tracks increases significantly as a 2055 function of $p_T$ , and also increases as the distance to the jet axis decreases, 2056 and the number of tracks in the jet increases (not shown). . . . .	58
2057	4.4 The light-jet efficiency of the low level tagger SV1 as a function of $b$ -jet 2058 efficiency for the nominal tracking setup (black) and for the case where fake 2059 tracks which are not from the decay of a $b$ -hadron are removed. The light-jet 2060 efficiency is decreased, demonstrating that the presence of fake tracks is 2061 detrimental to algorithm performance. . . . .	58

2062	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. . . . .	65
2069	5.2 The inputs to GN1 are the two jet features ( $n_{\text{jf}} = 2$ ), and an array of $n_{\text{tracks}}$ , where each track is described by 21 track features ( $n_{\text{tf}} = 21$ ). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1. . . . .	72
2073	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 rep- resentations are used to predict the jet flavour, the track origins, and the track-pair vertex compatibility. . . . .	72
2079	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. . . . .	77
2085	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 $20 < p_{\text{T}} < 250 \text{ GeV}$ . The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1 and 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. . . . .	79

2094	5.6 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet 2095 tagging efficiency for jets in the $Z'$ sample with $250 < p_T < 5000$ GeV. The 2096 ratio with respect to the performance of the DL1r algorithm is shown in the 2097 bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for 2098 DL1r and $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands are 2099 denoted by the shaded regions. At $b$ -jet tagging efficiencies less than $\sim 20\%$ , 2100 the light-jet rejection becomes so large that the effect of the low number of 2101 jets is visible. The lower $x$ -axis range is chosen to display the $b$ -jet tagging 2102 efficiencies usually probed in these regions of phase space. . . . .	80
2103	5.7 The $b$ -jet tagging efficiency for jets in the $t\bar{t}$ sample (left) and jets in the $Z'$ 2104 sample (right) as a function of jet $p_T$ with a fixed light-jet rejection of 100 in 2105 each bin. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and 2106 $f_c = 0.05$ is used for GN1 and GN1 Lep. Binomial error bands are denoted 2107 by the shaded regions. . . . .	81
2108	5.8 The $b$ -jet (left) and light-jet (right) rejections as a function of the $c$ -jet 2109 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV. The ratio to the 2110 performance of the DL1r algorithm is shown in the bottom panels. Binomial 2111 error bands are denoted by the shaded regions. At $c$ -jet tagging efficiencies 2112 than $\sim 25\%$ , the light-jet rejection becomes so large that the effect of the 2113 low number of jets is visible. The lower $x$ -axis range is chosen to display the 2114 $c$ -jet tagging efficiencies usually probed in these regions of phase space. . .	83
2115	5.9 The $b$ -jet (left) and light-jet (right) rejections as a function of the $c$ -jet 2116 tagging efficiency for $Z'$ jets with $250 < p_T < 5000$ GeV. The ratio to the 2117 performance of the DL1r algorithm is shown in the bottom panels. Binomial 2118 error bands are denoted by the shaded regions. The lower $x$ -axis range is 2119 chosen to display the $c$ -jet tagging efficiencies usually probed in these regions 2120 of phase space. . . . .	83

2121	5.10 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet 2122 tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the nominal 2123 GN1, in addition to configurations where no (GN1 No Aux), only the track 2124 classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives 2125 are deployed. The ratio to the performance of the DL1r algorithm is shown 2126 in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ 2127 for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted 2128 by the shaded regions. At $b$ -jet tagging efficiencies less than $\sim 65\%$ , the 2129 light-jet rejection become so large that the effect of the low number of jets 2130 are visible. The lower $x$ -axis range is chosen to display the $b$ -jet tagging 2131 efficiencies usually probed in these regions. . . . .	85
2132	5.11 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet 2133 tagging efficiency for $Z'$ jets with $250 < p_T < 5000$ GeV, for the nominal 2134 GN1, in addition to configurations where no (GN1 No Aux), only the track 2135 classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives 2136 are deployed. The ratio to the performance of the DL1r algorithm is shown 2137 in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ 2138 for DL1r and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted 2139 by the shaded regions. At $b$ -jet tagging efficiencies less than $\sim 25\%$ , the 2140 light-jet rejection become so large that the effect of the low number of jets 2141 are visible. The lower $x$ -axis range is chosen to display the $b$ -jet tagging 2142 efficiencies usually probed in these regions. . . . .	86
2143	5.12 Vertex finding efficiency as a function of jet $p_T$ for $b$ -jets in the $t\bar{t}$ sample 2144 using the exclusive (left) and inclusive (right) vertex finding approaches. 2145 Efficient vertex finding requires the recall of at least 65% of the tracks in 2146 the truth vertex, and allows no more than 50% of the tracks to be included 2147 incorrectly. Binomial error bands are denoted by the shaded regions. . . . .	89
2148	5.13 Inclusive vertex finding efficiency for multitrack truth vertices in $b$ -jets in 2149 the $t\bar{t}$ sample (left) and jets in the $Z'$ sample (right) as a function of jet $p_T$ . 2150 Efficient vertex finding requires the recall of at least 65% of the tracks in 2151 the truth vertex, and allows no more than 50% of the tracks to be included 2152 incorrectly. . . . .	90

2153	5.14 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. . . . .	92
2160	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 $\mu_R$ has been varied by a factor of 2 (“1up”) and 0.5 (“1down”). An exponential function has been fit to the ratio. . . . .	99
2165	6.2 Post-fit distributions for the 0-lepton (6.2a) and 2-lepton (6.2b) 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. . . . .	103

# 2168 List of tables

<small>2169</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>2170</small>	10	
<small>2171</small>					
<small>2172</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>2173</small>	<small>2174</small>	11
<small>2175</small>					
<small>2176</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 2.1.2). Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance. . . . .	<small>2177</small>	<small>2178</small>	24
<small>2179</small>					
<small>2180</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 $\Lambda^0$ decays, and hadronic interactions. . . . .	<small>2181</small>	<small>2182</small>	57
<small>2183</small>					
<small>2184</small>					
<small>2185</small>					
<small>2186</small>					
<small>2187</small>	4.2	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. . . . .	<small>2188</small>	<small>2189</small>	60
<small>2190</small>					
<small>2191</small>					
<small>2192</small>					
<small>2193</small>					

2194      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 $\theta$ 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. . . . .	61
2201      4.4	Good and fake track selection efficiencies for the combined $t\bar{t}$ and $Z'$ samples. Two working points are defined. . . . .	61
2203      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. . . . .	69
2210      5.2	Quality selections applied to tracks, where $d_0$ is the transverse IP of the track, $z_0$ is the longitudinal IP with respect to the PV and $\theta$ 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. . . . .	70
2217      5.3	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. . . . .	74
2220      5.4	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. . . . .	91

2227	6.1	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. . . . .	97
2228			
2229			
2230	6.2	V+jets acceptance uncertainties. $W + \text{jets}$ SR and CR uncertainties marked with a superscript † are correlated. The 1L $W + \text{jets}$ H/M uncertainty marked by * is applied as independent and uncorrelated NPs in both HP and LP signal regions. The 0L $W + \text{jets}$ Wbb Norm uncertainty is only applied when a floating normalisation for Wbb cannot be obtained from the 1L channel. A 30% uncertainty for Zbb norm is applied in the 1L channel when a floating normalisation for Zbb cannot be obtained from the 0L or 2L channels. . .	101
2231			
2232			
2233			
2234			
2235			
2236			