

# 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**

February 8, 2023

## **Declaration**

I, Samuel John Van Stroud confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Samuel Van Stroud

---

# Abstract

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

As  $b$ -jet identification relies on the successful reconstruction of charged particle trajectories (tracks), the tracking performance is investigated and potential improvements are assessed. Track reconstruction becomes increasingly difficult at high transverse momentum due to the increased multiplicity and collimation of tracks, and also due to the presence of displaced tracks that may occur as the result of the decay of a long-flying  $b$ -hadron. The investigations revealed that the track quality selections applied during track reconstruction are suboptimal for tracks inside high transverse momentum  $b$ -jets, motivating future studies into the optimisation of these cuts.

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

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

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

# Impact Statement

This thesis details research in experimental particle physics. The primary contributions are on the improvement of the data analysis algorithms which are used to process proton-proton collisions induced within the ATLAS detector at the Large Hadron Collider (LHC), and the analysis of candidate Higgs boson events.

The primary outcome of the research is an advancement of knowledge about how the Universe works on the most fundamental level, encoded for example in the improved measurement of the fundamental constants for the standard model, or in the observation of previously unseen particles or interactions. Although this kind of knowledge doesn't always have an immediate and direct relevance for society, potential applications are impossible to rule out and could have a very large impact further in the future, as has been seen with previous advancements in fundamental science.

The research does find indirect application in the form of associated technological developments that have transferable application within different fields. The cutting-edge techniques developed at CERN for ATLAS and the LHC have found many spin-off applications elsewhere in society, for example the World Wide Web, high-field magnet technology in MRI, touch-screen technology and cloud computing. Fundamental physics, as a proposer of novel and difficult problems, can therefore be seen as a way to generate ideas for new technologies.

Working in the field also helps to train skilled researchers, which can be redeployed to other areas of society to tackle various problems. In this thesis advanced statistical and data science methods are

deployed. Such methods currently find wide and varied use in many fields.

Finally, the work carried at ATLAS and the LHC is widely publicised – support of and interest in fundamental physics research helps to generate excitement about science and technology, and educate people about how the Universe works. This in turn attracts people into the area, propagating the benefits described above.

## Acknowledgements

Firstly I give thanks to my supervisor Tim Scanlon for all the guidance and support he has offered over the course of this doctorate. Tim has always been consistent with clear explanations and sound advice throughout the last four years. I would also like to thank everyone I've worked with at ATLAS and at UCL. In particular I have Jonathan Shlomi to thank for the fruitful collaboration on advancements in flavour tagging. I would also like to thank Brian Moser and Hannah Arnold for their patient support during the course of the  $VH$ ,  $H \rightarrow b\bar{b}$  analysis. I'm grateful to everyone I've worked with in the ATLAS Flavour Tagging and Tracking groups, in particular to Gabriel Facini, Valerio Dao, Bingxuan Liu and Francesco Di Bello for their guidance, and Dan Guest for his fastidious merge request reviews. This thesis was made in L<sup>A</sup>T<sub>E</sub>X 2 <sub>$\varepsilon$</sub>  using the “heptesis” class [1].

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                     | <b>12</b> |
| <b>2</b> | <b>Theoretical Framework</b>                            | <b>14</b> |
| 2.1      | The Standard Model . . . . .                            | 14        |
| 2.1.1    | Quantum Electrodynamics . . . . .                       | 15        |
| 2.1.2    | Quantum Chromodynamics . . . . .                        | 17        |
| 2.1.3    | The Electroweak Sector . . . . .                        | 18        |
| 2.2      | The Higgs Mechanism . . . . .                           | 19        |
| 2.2.1    | Electroweak Symmetry Breaking . . . . .                 | 20        |
| 2.2.2    | Fermionic Yukawa Coupling . . . . .                     | 23        |
| 2.2.3    | Higgs Sector Phenomenology . . . . .                    | 25        |
| <b>3</b> | <b>The Large Hadron Collider and the ATLAS Detector</b> | <b>28</b> |
| 3.1      | The Large Hadron Collider . . . . .                     | 29        |
| 3.2      | Coordinate System & Collider Definitions . . . . .      | 30        |
| 3.2.1    | ATLAS Coordinate System . . . . .                       | 31        |
| 3.2.2    | Track Parameterisation . . . . .                        | 32        |
| 3.2.3    | Hadron Collider Definitions . . . . .                   | 32        |
| 3.3      | The ATLAS Detector . . . . .                            | 36        |
| 3.3.1    | Inner Detector . . . . .                                | 37        |
| 3.3.2    | Calorimeters . . . . .                                  | 40        |
| 3.3.3    | Muon Spectrometer . . . . .                             | 42        |
| 3.3.4    | The Trigger . . . . .                                   | 43        |
| 3.4      | Reconstructed Physics Objects . . . . .                 | 44        |
| 3.4.1    | Tracks . . . . .  | 44        |
| 3.4.2    | Vertices . . . . .                                      | 47        |
| 3.4.3    | Jets . . . . .  | 48        |
| 3.4.4    | Leptons . . . . .                                       | 50        |

|          |  |           |
|----------|--|-----------|
| 3.4.5    | Missing Transverse Momentum . . . . .                              | 52        |
| <b>4</b> | <b>Tracking and flavour tagging</b>                                | <b>53</b> |
| 4.1      | <i>b</i> -hadron Reconstruction . . . . .                          | 54        |
| 4.1.1    | Decay Topology . . . . .   | 54        |
| 4.1.2    | Challenges Facing <i>b</i> -hadron Reconstruction . . . . .        | 56        |
| 4.2      | Investigations into High $p_T$ <i>b</i> -hadron Tracking . . . . . | 60        |
| 4.2.1    | Shared Hits . . . . .  | 60        |
| 4.2.2    | Global $\chi^2$ Fitter Outlier Removal . . . . .                   | 61        |
| 4.3      | Conclusion . . . . .   | 63        |
| <b>5</b> | <b>Track Classification MVA</b>                                    | <b>65</b> |
| 5.1      | Machine Learning Background . . . . .                              | 65        |
| 5.1.1    | Neural Networks . . . . .  | 66        |
| 5.1.2    | Training with Gradient Descent . . . . .                           | 68        |
| 5.2      | Graph Neural Network Theory . . . . .                              | 69        |
| 5.3      | Track Truth Origin Labelling . . . . .                             | 70        |
| 5.4      | Fake Track Identification Tool . . . . .                           | 72        |
| 5.4.1    | Datasets . . . . .   | 72        |
| 5.4.2    | Model Inputs . . . . .   | 74        |
| 5.4.3    | Model Hyperparameters . . . . .                                    | 76        |
| 5.4.4    | Results . . . . .  | 76        |
| 5.5      | <i>b</i> -hadron Track Identification . . . . .                    | 78        |
| 5.6      | Combined Approach . . . . .  | 79        |
| 5.7      | Conclusion . . . . .   | 80        |
| <b>6</b> | <b>Graph Neural Network Flavour Tagger</b>                         | <b>82</b> |
| 6.1      | Motivation . . . . .   | 82        |
| 6.2      | Experiemental Setup . . . . .                                      | 85        |
| 6.2.1    | Datasets . . . . .   | 85        |
| 6.3      | Model Architecture . . . . .                                       | 85        |
| 6.3.1    | Model Inputs . . . . .   | 85        |
| 6.3.2    | Auxiliary Training Objectives . . . . .                            | 86        |
| 6.3.3    | Architecture . . . . .   | 89        |
| 6.3.4    | Training . . . . .   | 93        |

|          |   |            |
|----------|---|------------|
| 6.4      | Results . . . . .   | 94         |
| 6.4.1    | <i>b</i> -tagging Performance . . . . .                   | 95         |
| 6.4.2    | <i>c</i> -tagging Performance . . . . .                   | 97         |
| 6.4.3    | Ablations . . . . .                                       | 100        |
| 6.4.4    | Inclusion of Low-Level Vertexing Algorithms . . . . .     | 102        |
| 6.4.5    | Jet Display Diagrams . . . . .                            | 102        |
| 6.4.6    | Vertexing Performance . . . . .                           | 104        |
| 6.4.7    | Track Classification Performance . . . . .                | 110        |
| 6.4.8    | Looser Track Selection . . . . .                          | 112        |
| 6.5      | Other Implementations of GN1 . . . . .                    | 113        |
| 6.6      | Conclusion . . . . .                                      | 115        |
| <b>7</b> | <b>Boosted VHbb Analysis</b>                              | <b>119</b> |
| 7.1      | Analysis Overview . . . . .                               | 120        |
| 7.1.1    | Data & Simulated Samples . . . . .                        | 122        |
| 7.1.2    | Object Reconstruction . . . . .                           | 122        |
| 7.1.3    | Selection Criteria . . . . .                              | 124        |
| 7.1.4    | Control Regions . . . . .                                 | 125        |
| 7.1.5    | Background Composition . . . . .                          | 125        |
| 7.2      | Systematic Uncertainties & Background Modelling . . . . . | 128        |
| 7.2.1    | Implementation of Variations . . . . .                    | 129        |
| 7.2.2    | Sources of Systematic Uncertainties . . . . .             | 130        |
| 7.2.3    | Vector Boson + Jets Modelling . . . . .                   | 132        |
| 7.2.4    | Diboson Modelling . . . . .                               | 136        |
| 7.3      | Statistical Treatment . . . . .                           | 141        |
| 7.3.1    | Likelihood Function . . . . .                             | 142        |
| 7.3.2    | Background Normalisations . . . . .                       | 144        |
| 7.3.3    | Asimov Dataset & Expected Results . . . . .               | 144        |
| 7.4      | Results . . . . .   | 145        |
| 7.4.1    | Post-fit Distributions . . . . .                          | 145        |
| 7.4.2    | Observed Signal Strength & Significance . . . . .         | 146        |
| 7.4.3    | Impact of Systematics . . . . .                           | 150        |
| 7.5      | Conclusion . . . . .                                      | 154        |
| <b>8</b> | <b>Conclusion</b>   | <b>155</b> |
| 8.1      | Summary . . . . .   | 155        |

---

**Contents** **11**

|                           |            |
|---------------------------|------------|
| 8.2 Future Work . . . . . | 156        |
| <b>Bibliography</b>       | <b>157</b> |
| <b>List of figures</b>    | <b>171</b> |
| <b>List of tables</b>     | <b>183</b> |

# <sup>2</sup> Chapter 1

## <sup>3</sup> Introduction

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

- 21 The author's contribution to the work presented in this thesis is as follows.
- 22 **Tracking:** The author was been an active member of the Cluster and Tracking  
23 in Dense Environments group for the duration of their qualification task on the  
24 understanding of tracking performance at high transverse momentum. The author  
25 played a key role in software r22 validation studies for the tracking group, including  
26 the validation of the quasi-stable particle interaction simulation and the radiation  
27 damage Monte-Carlo simulation. The author helped design and improve several  
28 tracking software frameworks, and contributed to heavy flavour tracking efficiency  
29 studies in dense environments.
- 30 ***b*-tagging:** The author has been an active member of the Flavour Tagging group  
31 since September 2014. The author played a key role in investigating the performance  
32 of the low level taggers at high transverse momentum and led studies into the  
33 labelling and classification of track origins. Based on work by Jonathan Shlomi,  
34 the author helped develop a new flavour tagging algorithm which offers a large  
35 performance improvement with respect to the current state of the art. The author  
36 was the primary editor of a public note associated with this work [2], and also  
37 contributed to the proliferation of GN1 to the trigger, High Luminosity LHC, and  
38  $X \rightarrow bb$  use cases. The author also played a key role in software r22 validation  
39 studies for the Flavour Tagging group, including the validation of the quasi-stable  
40 particle interaction simulation. The author maintains and contributes to various  
41 software frameworks used in the Flavour Tagging group, and contributes to group  
42 documentation.
- 43 **Higgs:** The author was an active member of the Boosted VHbb analysis group. The  
44 author performed various studies deriving systematic uncertainties for the  $V+jets$   
45 and diboson backgrounds. The author also produced and maintained samples, ran fit  
46 studies and cross checks, and gave the diboson unblinding approval talk to the Higgs  
47 group. The author also contributed to the developement of the analysis software.

<sup>48</sup> Chapter 2

<sup>49</sup> Theoretical Framework

<sup>50</sup> The Standard Model (SM) of particle physics is the theory describing all known  
<sup>51</sup> elementary particles and their interactions via three of the four fundamental forces.  
<sup>52</sup> Developed by merging the successful theories of quantum mechanics and relativity  
<sup>53</sup> in the second half of the 20th century, the SM's position today at the centre of our  
<sup>54</sup> understanding of the nature of the Universe is firmly established by an unparalleled  
<sup>55</sup> level of agreement between the predictions from the model and experimental results  
<sup>56</sup> [3, 4].

<sup>57</sup> The SM has predicted the discovery of the top and bottom quarks [5–7], the  $W$   
<sup>58</sup> and  $Z$  bosons [8], and the tau neutrino [9]. The last missing piece of the SM to be  
<sup>59</sup> discovered was the Higgs boson, first theorised in the 1960s [10–12], and eventually  
<sup>60</sup> observed at the LHC in 2012 [13, 14]. After its discovery, much ongoing work has  
<sup>61</sup> been carried out performing detailed measurements of its mass and interactions with  
<sup>62</sup> other particles.

<sup>63</sup> In this chapter, an overview of the SM is given in Section 2.1, and a more detailed  
<sup>64</sup> discussion of the Higgs sector and Higgs phenomenology is provided in Section 2.2.

<sup>65</sup> 2.1 The Standard Model

<sup>66</sup> The SM is formulated in the language of Quantum Field Theory (QFT). In this  
<sup>67</sup> framework, particles are localised excitations of corresponding quantum fields, which  
<sup>68</sup> are operator-valued distributions across spacetime.

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

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

The SM provides a mathematical description of how three of the four fundamental forces interact with the matter content of the Universe. The SM contains 12 spin-1/2 fermions, listed in Table 2.1, and five bosons listed in Table 2.2.

| <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               | $2/3$             |
|                   | $\nu_e$        | $< 1.1 \times 10^{-6}$ | 0                 | $d$            | 4.67               | $-1/3$            |
| Second            | $\mu$          | 105.7                  | -1                | $c$            | $1.27 \times 10^3$ | $2/3$             |
|                   | $\nu_\mu$      | $< 0.19$               | 0                 | $s$            | 93.4               | $-1/3$            |
| Third             | $\tau$         | 1776.9                 | -1                | $t$            | $173 \times 10^3$  | $2/3$             |
|                   | $\nu_\tau$     | $< 18.2$               | 0                 | $b$            | $4.18 \times 10^3$ | $-1/3$            |

**Table 2.1:** The fermions of the SM [15]. Three generations of particles are present. Also present (unlisted) are the antiparticles, which are identical to the particles up to a reversed charge sign.

80

### 2.1.1 Quantum Electrodynamics

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

| Name               | Symbol   | Mass [GeV]            | Charge [ $e$ ]        | Spin |
|--------------------|----------|-----------------------|-----------------------|------|
| Photon             | $\gamma$ | $< 1 \times 10^{-27}$ | $< 1 \times 10^{-46}$ | 1    |
| Charged Weak boson | $W^\pm$  | $80.377 \pm 0.012$    | $\pm 1$               | 1    |
| Neutral 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 2.2:** The bosons of the SM [15]. The photon, weak bosons and gluons are gauge bosons arising from gauge symmetries, and carry the four fundamental forces of the SM. The recently discovered Higgs boson is the only fundamental scalar particle in the SM.

<sup>86</sup> Dirac Lagrangian density is

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

<sup>87</sup> where  $\cancel{\partial} = \gamma^\mu \partial_\mu$  denotes the contraction with the Dirac gamma matrices  $\gamma^\mu$  (summation over up-down pairs of indices is assumed). Application of the Euler-Lagrange equation on Eq. (2.2) yields the Dirac equation

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

<sup>90</sup> Suppose some fundamental symmetry that requires invariance under a local  $U(1)$   
<sup>91</sup> gauge transformation

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

<sup>92</sup> where  $\alpha$  varies over every spacetime point  $x$ . Under this transformation, the Dirac  
<sup>93</sup> equation transforms as

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

<sup>94</sup> For the Dirac equation to remain invariant under the transformation in Eq. (2.4),  
<sup>95</sup> a new field  $A_\mu$  which transforms as  $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \alpha(x)$  must be added. The  
<sup>96</sup> transformed interaction term

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

<sup>97</sup> will then cancel the asymmetric term in Eq. (2.5) as required. The  $U(1)$  invariant  
<sup>98</sup> Lagrangain can therefore be constructed by adding an interaction between  $\psi$  and  
<sup>99</sup>  $A_\mu$  to Eq. (2.2). For completeness, the kinetic term for the new field  $A_\mu$  is  
<sup>100</sup> also added in terms of  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ , which is trivially invariant under the  
<sup>101</sup> transformation in Eq. (2.4). The interaction term is typically absorbed into the  
<sup>102</sup> covariant derivative  $D_\mu = \partial_\mu + iqA_\mu$ , thus named as it transforms in the same way as  
<sup>103</sup> the field  $\psi$ . Collecting these modifications to Eq. (2.2) yields the QED Lagrangain

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

<sup>104</sup> The quadratic term  $A_\mu A^\mu$  is not invariant and therefore the field  $A_\mu$  must be  
<sup>105</sup> massless. Requiring invariance under local  $U(1)$  gauge transformations necessitated  
<sup>106</sup> the addition of a new field  $A_\mu$ , interpreted as the photon field, which interacts with  
<sup>107</sup> charged matter. In the SM, the QED Lagrangian is absorbed into the electroweak  
<sup>108</sup> sector, discussed in Section 2.1.3.

### <sup>109</sup> 2.1.2 Quantum Chromodynamics

<sup>110</sup> Quantum Chromodynamics (QCD) is the study of quarks, gluons and their interactions.  
<sup>111</sup> Quarks and gluons carry colour charge, which comes in three kinds, called  
<sup>112</sup> red, green and blue. While the  $U(1)$  symmetry group in Section 2.1.1 was Abelian,  
<sup>113</sup> the QCD Lagrangian is specified by requiring invariance under transformations from  
<sup>114</sup> the non-Abelian  $SU(3)$  group, making it a Yang-Mills theory [16] which requires the  
<sup>115</sup> addition of self-interacting gauge fields. The infinitesimal  $SU(3)$  group generators  
<sup>116</sup> are given by  $T_a = \lambda_a/2$ , where  $\lambda_a$  are the eight Gell-Mann matrices. These span the  
<sup>117</sup> space of infinitesimal group transformations and do not commute with each other,  
<sup>118</sup> instead satisfying the commutation relation

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

<sup>119</sup> where  $f_{abc}$  are the group's structure constants. Consider the six quark fields  $q_k = q_k(x)$ .  
<sup>120</sup> Each flavour of quark  $q_k$  transforms in the fundamental triplet representation, in  
<sup>121</sup> which each component of the triplet corresponds to the colour quantum number  
<sup>122</sup> for red, green and blue colour charged respectively.  $G_{\mu\nu}^a$  are the eight gluon field

123 strength tensors, one for each generator  $T_a$ , defined as

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

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

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

126 The full QCD Lagrangian is then given by

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

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

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

### 137 2.1.3 The Electroweak Sector

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

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

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

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

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

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

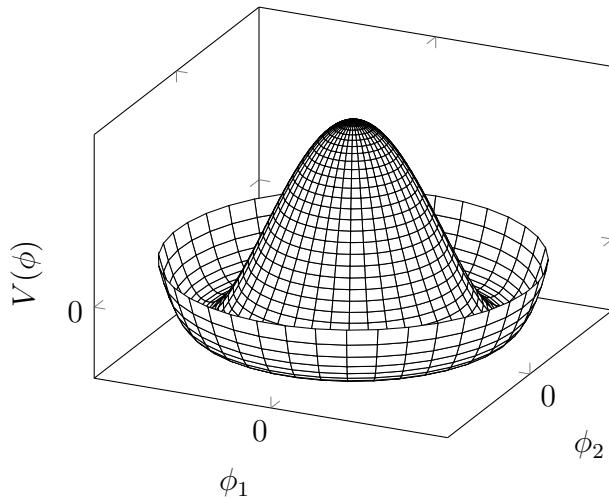
- 153 The weak force violates parity conservation [20–22], i.e. invariance under parity  
154 transformations (mirror reflections). Only left handed fermions participate in the  
155 weak interaction. Since there is no other force through which neutrinos interact with  
156 other particles, there are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

## 157 2.2 The Higgs Mechanism

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

### <sup>168</sup> 2.2.1 Electroweak Symmetry Breaking

<sup>169</sup> Spontaneous symmetry breaking (SSB) is a key part of the Higgs mechanism. It  
<sup>170</sup> is the transition of a physical system from a state of manifest symmetry to a state  
<sup>171</sup> of hidden, or *broken*, symmetry. In particular, this applies to physical systems  
<sup>172</sup> where the Lagrangian observes some symmetry, but the lowest energy vacuum states  
<sup>173</sup> do not exhibit that same symmetry. In other words, the symmetry is broken for  
<sup>174</sup> perturbations around the vacuum state.



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

<sup>175</sup> Consider gauge fields from the local  $SU(2)_L \otimes U(1)_Y$  symmetry group discussed in  
<sup>176</sup> Section 2.1.3 coupled to a complex scalar field  $\phi = \phi(x)$ . The scalar field  $\phi$  transforms  
<sup>177</sup> as a weak isospin doublet. Omitting the kinetic term of the gauge fields, and writing  
<sup>178</sup>  $\phi^2 \equiv \phi^\dagger \phi$ , the Lagrangian is

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

<sup>179</sup> where the covariant derivative is given by

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

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

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

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

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

$$\phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v + H \end{bmatrix}, \quad (2.17)$$

where  $H$  is a real scalar field, the *true vacuum* Higgs field. Substituting this into Eq. (2.14) and identifying physical fields from the quadratic terms of linear combinations of unphysical fields, one can write the physical fields  $W_\mu^\pm$ ,  $Z_\mu$  and  $A_\mu$

206 in terms of the original fields  $A_\mu^a$  and  $B_\mu$ . This gives

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

207 where  $\theta_W$  is the weak mixing angle defined by

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

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

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

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

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

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

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

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

### 232 2.2.2 Fermionic Yukawa Coupling

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

239 The Higgs field  $\phi$  transforms as an  $SU(2)$  doublet with  $Y = 1/2$ , as does the left-  
240 handed fermion field  $\psi_L$ . The right-handed fermion field  $\psi_R$  transforms as an  $SU(2)$   
241 singlet.

### 242 Lepton Masses

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

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

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

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

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

254 **Quark Masses**

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

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

259 and transforms in the same way as  $\phi$ . This field can be used to write an additional  
260 Yukawa coupling which provides mass for the up-type quarks in a similar way as  
261 before.

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

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

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

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

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

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

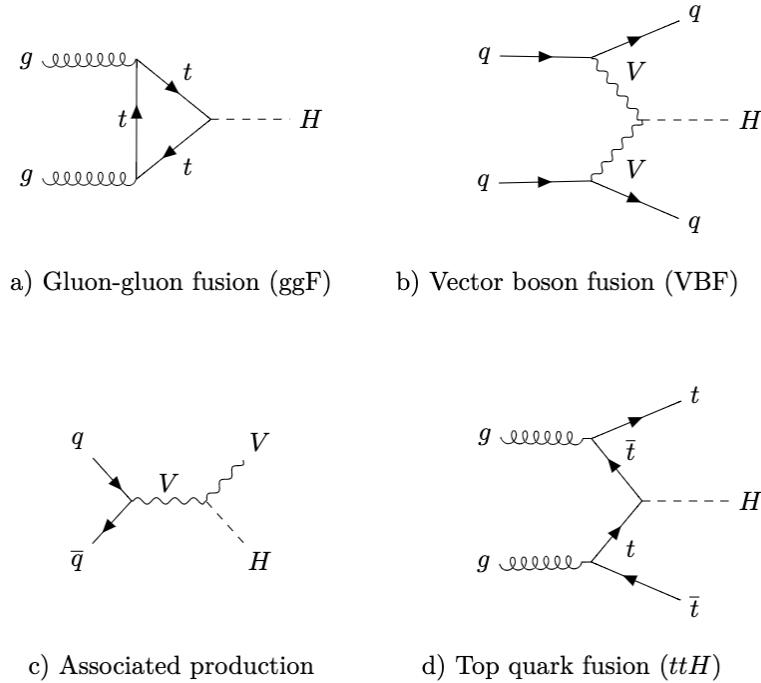
### <sup>274</sup> 2.2.3 Higgs Sector Phenomenology

<sup>275</sup> As previous discussed in this section, the Higgs field plays a key role in the SM,  
<sup>276</sup> giving mass to fundamental particles. The strength of the coupling between the  
<sup>277</sup> Higgs field and another particle is proportional to that particle's mass. This fact  
<sup>278</sup> dictates which production mechanisms and decay modes are dominant at the LHC.  
<sup>279</sup> The cross sections for different production mechanisms at a centre of mass energy  
<sup>280</sup>  $\sqrt{s} = 13 \text{ TeV}$  are shown as a function of the Higgs mass  $m_H$  in Fig. 2.3. Higgs boson  
<sup>281</sup> production occurs mainly through four modes, shown in Fig. 2.2. The dominant  
<sup>282</sup> production mode is gluon-gluon fusion ( $pp \rightarrow H$ ), which is predominantly mediated  
<sup>283</sup> by a virtual top quark loop. Vector boson fusion ( $pp \rightarrow qqH$ ) is the second most  
<sup>284</sup> dominant production mechanism, in which a pair of  $W$  or  $Z$  bosons fuse to produce  
<sup>285</sup> a Higgs after being radiated by two quarks, which also occur in the final state. Next  
<sup>286</sup> most common is the associated production of a Higgs boson and a vector boson  
<sup>287</sup> ( $pp \rightarrow VH$ ), in which a pair of quarks fuse to produce a single  $W$  or  $Z$  boson which  
<sup>288</sup> radiates a Higgs. The final of the four leading production modes is top quark fusion,  
<sup>289</sup> in which two gluons each radiate a quark-antiquark pair, and a quark from each pair  
<sup>290</sup> fuses to produce a Higgs boson.

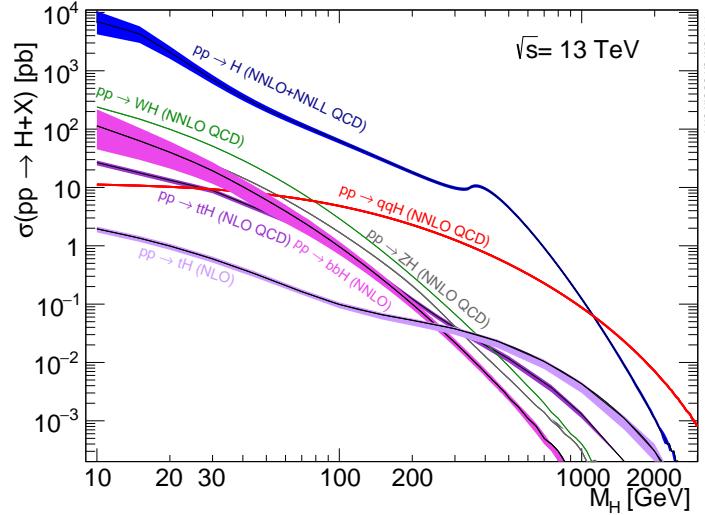
<sup>291</sup> Although gluon-gluon fusion is the dominant production mode, for hadronic decays  
<sup>292</sup> of the Higgs boson the associated production with a vector boson has the advantage  
<sup>293</sup> of leading to a more conspicuous final state due to the likelihood of the vector bosons  
<sup>294</sup> decaying leptons. Leptons provide a clean signals to detect and trigger on.

<sup>295</sup> Since the Higgs boson couples proportional to mass as already mentioned, decays  
<sup>296</sup> to heavier particles are favoured. The branching ratios of different Higgs boson  
<sup>297</sup> decay modes are shown as a function of  $m_H$  in Fig. 2.4. Approximately 58% of the  
<sup>298</sup> time the Higgs boson decays to a pair of  $b$ -quarks, the dominant decay mode. The  
<sup>299</sup> next heaviest fermions are the tau lepton and the  $c$ -quark, decays to pairs of these  
<sup>300</sup> particles happen approximately an order of magnitude less often. Decays to pairs  
<sup>301</sup> of vector bosons are via a virtual off shell Higgs boson only. While the  $H \rightarrow \gamma\gamma$  and  
<sup>302</sup>  $H \rightarrow ZZ$  branching ratios are small compared with fermionic decay modes (around  
<sup>303</sup> 0.2% for  $H \rightarrow \gamma\gamma$ ), these decay channels were instrumental in the initial discovery of  
<sup>304</sup> the Higgs due to the low level of background processes which mimic the final state.

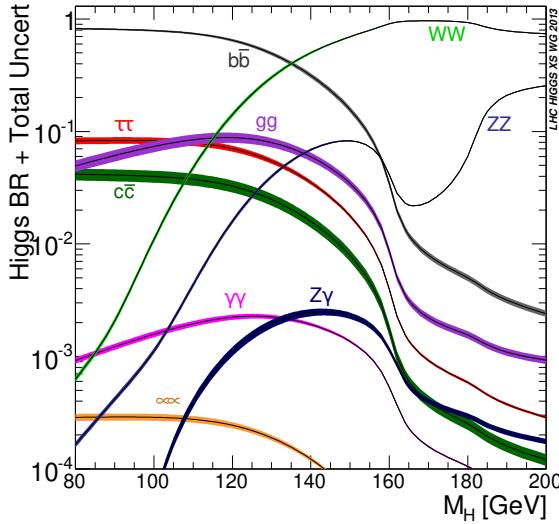
<sup>305</sup> This thesis presents a measurement of the Higgs bosons production rate using events  
<sup>306</sup> with a Higgs boson produced in association with vector boson and decaying to a pair



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



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



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

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

<sub>312</sub> Chapter 3

<sub>313</sub> The Large Hadron Collider and the  
<sub>314</sub> ATLAS Detector

<sub>315</sub> Since the completion of its construction in 2008, the Large Hadron Collider (LHC) [28]  
<sub>316</sub> at CERN has extended the frontiers of particle physics through its unprecedented  
<sub>317</sub> energy and luminosity. The LHC accelerates protons around a 27 km ring until they  
<sub>318</sub> are travelling just  $3 \text{ m s}^{-1}$  slower than the speed of light, at which point they  
<sub>319</sub> are made to collide. The protons travel round the ring 11,000 times per second in  
<sub>320</sub> two concentric beams, which are guided by superconducting magnets cooled using  
<sub>321</sub> liquid helium to  $-271.3^\circ\text{C}$  ( $1.9\text{ K}$ ). The beams travel in opposite directions around  
<sub>322</sub> the ring and are crossed at four locations so that collisions between protons can  
<sub>323</sub> take place. Around these collision points four specialised detectors, ALICE [29],  
<sub>324</sub> CMS [30], LHCb [31] and ATLAS [32], are located to capture information about the  
<sub>325</sub> products of the collisions.

<sub>326</sub> In this chapter, a brief overview of the LHC and the accelerator complex at CERN  
<sub>327</sub> is given in Section 3.1. The coordinate system used at the ATLAS detector and  
<sub>328</sub> other common definitions are introduced in Section 3.2. Next, an overview of the  
<sub>329</sub> different detector systems is provided in Section 3.3, and finally descriptions of  
<sub>330</sub> various commonly used reconstructed objects is given in Section 3.4.

### <sup>331</sup> 3.1 The Large Hadron Collider

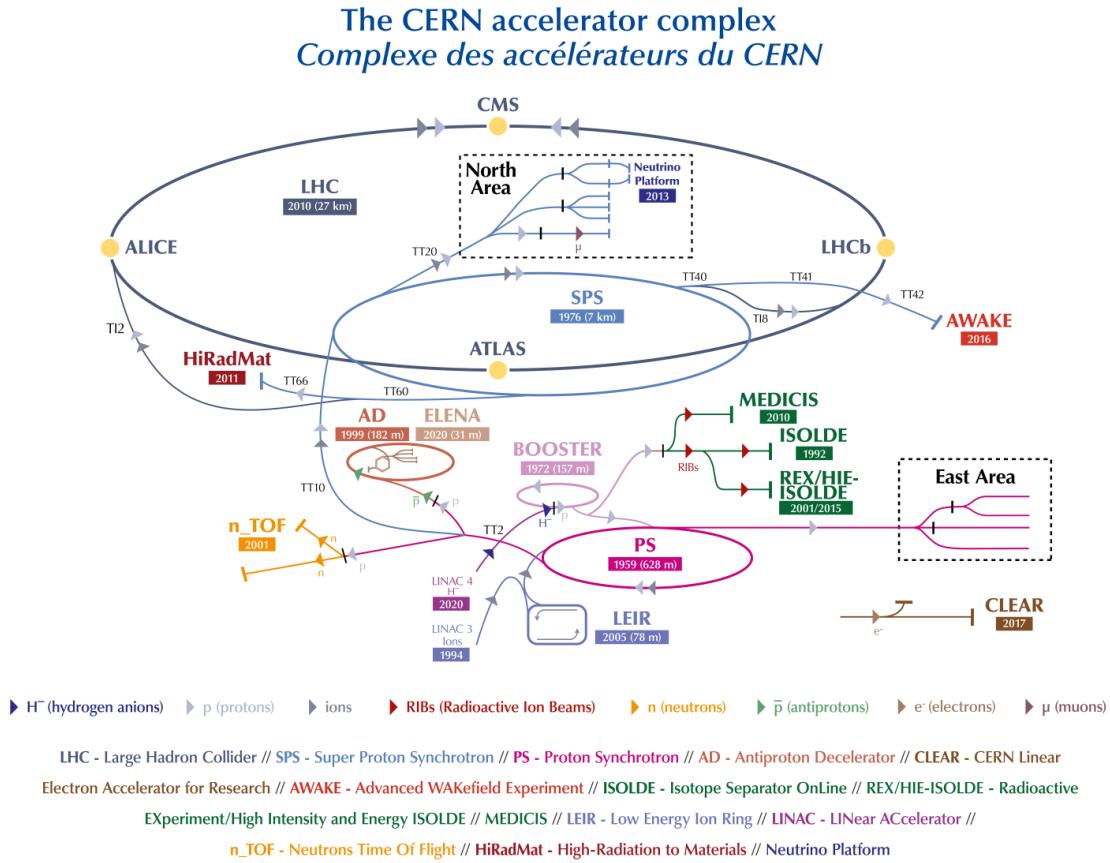
<sup>332</sup> The LHC is operated in multi-year *runs* during which beams of protons are circulated  
<sup>333</sup> and collided. Between runs there are periods of shutdown while the accelerator and  
<sup>334</sup> detector machinery is maintained and upgraded. Run 1 began in 2010 when the LHC  
<sup>335</sup> collided proton bunches, each containing more than  $10^{11}$  particles, 20 million times  
<sup>336</sup> per second, providing 7 TeV proton-proton collisions at instantaneous luminosities  
<sup>337</sup> of up to  $2.1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ . The centre-of-mass energy was increased to 8 TeV  
<sup>338</sup> towards the end of Run 1 in 2012. Run 2, which spanned in 2015–2018, further  
<sup>339</sup> increased the the proton-proton collision energy to 13 TeV. During Run 2 the bunch  
<sup>340</sup> spacing was reduced, leading to a collisison rate of 40 MHz. Over the course of  
<sup>341</sup> Run 2 a total usable integrated luminosity of  $139 \text{ fb}^{-1}$  was recorded. 2022 marked the  
<sup>342</sup> beginning of Run 3 which, with a higher center of mass energy and peak luminosity,  
<sup>343</sup> is expected to culminate in the approximate tripling of the dataset size. A summary  
<sup>344</sup> of key information about each run is listed in Table 3.1.

| Period | Year      | $\sqrt{s}$ [TeV] | $\langle\mu\rangle$ | Bunch spacing [ns] | Luminosity [ $\text{cm}^{-2} \text{ s}^{-1}$ ] |
|--------|-----------|------------------|---------------------|--------------------|--|
| Run 1  | 2010–2012 | 7–8              | 18                  | 50                 | $8 \times 10^{33}$                             |
| Run 2  | 2015–2018 | 13               | 34                  | 25                 | $1\text{--}2 \times 10^{34}$                   |
| Run 3  | 2022–2025 | 13.6             | 50                  | 25                 | $2 \times 10^{34}$                             |

**Table 3.1:** Overview of the different LHC runs [33,34]. The average number of interactions per bunch-crossing is denoted as  $\langle\mu\rangle$  (see Section 3.2.3), and is here averaged over the entire run. Numbers for Run 3 are preliminary and are only provided to give an indication of expected performance.

<sup>345</sup> An overview of the accelerator complex at CERN is shown in Fig. 3.1. The LHC is  
<sup>346</sup> at the final stage of a chain of accelerators which incrementally step-up the energy  
<sup>347</sup> of incoming protons. The first accelerator is Linac4, a linear accelerator which  
<sup>348</sup> accelerates negative hydrogen ions to an energy of 160 MeV. Upon leaving Linac4,  
<sup>349</sup> the ions are stripped of both electrons and the resulting protons are fed into the  
<sup>350</sup> Proton Synchrotron Booster (PSB), which increases the energy of the protons to  
<sup>351</sup> 2 GeV. The protons leaving the PSB are passed to the Proton Synchrotron (PS),  
<sup>352</sup> which increases the energy to 26 GeV, and then from the PS to the Super Proton  
<sup>353</sup> Synchrotron (SPS) which further increases the energy to 450 GeV. Finally, the proton

354 beams are injected in the LHC where they are accelerated to their final energy of  
 355 6.5 TeV (for Run 2).



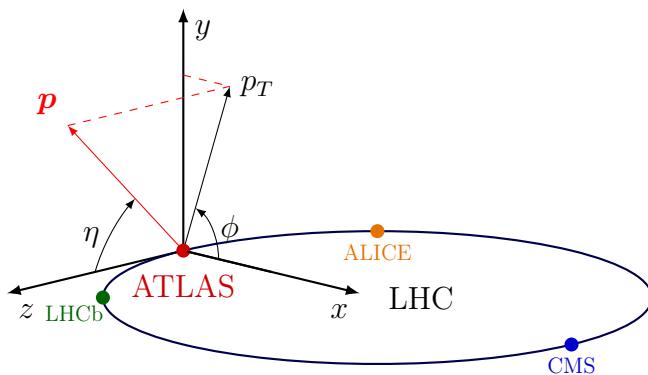
**Figure 3.1:** An overview of the CERN accelerator complex [35]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC.

## 3.2 Coordinate System & Collider Definitions

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

### 361 3.2.1 ATLAS Coordinate System

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



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

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

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

370 The pseudorapidity is a convenient quantity to work with as differences in  $\eta$  are  
 371 invariant under Lorentz boosts. In addition, particle production is constant as a  
 372 function of  $\eta$ .

373 The transverse momentum  $p_T$  of an object is the sum in quadrature of the momenta  
 374 in the transverse plane

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

<sup>375</sup> Angular distance between two objects is measured in units of  $\Delta R$  and is defined as  
<sup>376</sup> the sum in quadrature of the  $\eta$  and  $\phi$  displacements

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

### <sup>377</sup> 3.2.2 Track Parameterisation

<sup>378</sup> The trajectories of charged particle tracks are parameterised as a helix which is  
<sup>379</sup> fully specified using five parameters:  $(d_0, z_0, \phi, \theta, q/p)$ . Transverse and longitudinal  
<sup>380</sup> impact parameters (IP)  $d_0$  and  $z_0$  specify the closest approach of the trajectory of  
<sup>381</sup> a particle to the origin. The transverse IP  $d_0$  and longitudinal IP  $z_0$  are measured  
<sup>382</sup> with respect to the hard scatter primary vertex (see Section 3.4.2).  $\phi$  and  $\theta$  are  
<sup>383</sup> the azimuthal and polar angles respectively, and  $q/p$  is the measured charge on the  
<sup>384</sup> track<sup>1</sup> divided by the scalar 3-momentum. Fig. 3.3 shows each of these parameters  
<sup>385</sup> diagrammatically.

<sup>386</sup> Impact parameter significances are defined as the IP divided by its corresponding  
<sup>387</sup> uncertainty,  $s(d_0) = d_0/\sigma(d_0)$  and  $s(z_0) = z_0/\sigma(z_0)$ . When used in flavour tagging  
<sup>388</sup> (see Chapter 4), track IP significances are lifetime signed according to the track's  
<sup>389</sup> direction with respect to the jet axis and the primary vertex [38]. The sign IP  
<sup>390</sup> significances is positive if the track crosses the jet axis in front of the primary vertex  
<sup>391</sup> and negative if the crossing is behind the primary vertex.

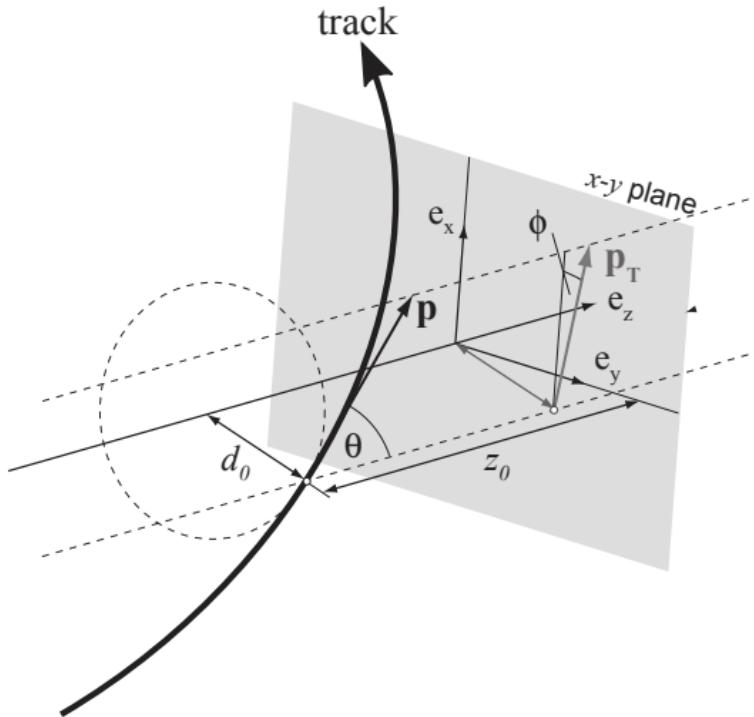
### <sup>392</sup> 3.2.3 Hadron Collider Definitions

#### <sup>393</sup> Cross Section

<sup>394</sup> The cross section  $\sigma$  is closely related to the probability of an interaction between  
<sup>395</sup> two colliding particles, and is analogous to an effective cross-sectional area of the  
<sup>396</sup> particles. The cross section of a process depends on the transition matrix element,  
<sup>397</sup> obtained using the Feynman rules of the theory which are derived using QFT, and a  
<sup>398</sup> phase space integral. At hadron colliders such as the LHC, the proton-proton cross

---

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



**Figure 3.3:** The track parameterisation used at the ATLAS detector. Five coordinates ( $d_0, z_0, \phi, \theta, q/p$ ) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum  $\mathbf{p}$  and the transverse momentum  $p_T$  (defined in Eq. (3.2)). The basis vectors  $e_x, e_y$  and  $e_z$  are also shown. Reproduced from Ref. [37].

<sup>399</sup> section can be factorised as

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

<sup>400</sup> The partonic cross section can be calculated at sufficiently high energies such as  
<sup>401</sup> those found at the LHC, while the parton distribution functions (PDFs) have to be  
<sup>402</sup> extracted from experimental results.

### <sup>403</sup> Luminosity

<sup>404</sup> The total number of proton-proton collisions  $N$  is related to the total  $pp$  cross  $\sigma$   
<sup>405</sup> section by the integrated luminosity  $L$ , as in

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

<sup>406</sup> The instantaneous luminosity  $\mathcal{L}$  relates the cross section to the number of collisions  
<sup>407</sup> per unit time. For two colliding bunched proton beams, it is defined as

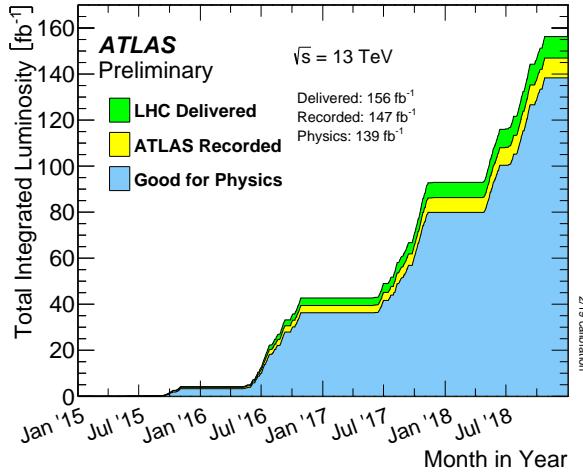
$$\mathcal{L} = \frac{1}{\sigma} \frac{dN}{dt} = \frac{fn_1 n_2}{4\pi\sigma_x\sigma_y}, \quad (3.6)$$

<sup>408</sup> where  $n_1$  and  $n_2$  are the number of protons in the colliding bunches,  $f$  is the bunch  
<sup>409</sup> crossing frequency, and  $\sigma_x$  and  $\sigma_y$  are the rms width of the beam in the horizontal  
<sup>410</sup> and vertical directions.

<sup>411</sup> The total luminosity recorded over the course of Run 2 is shown in Fig. 3.4. In  
<sup>412</sup> total,  $139 \text{ fb}^{-1}$  of usable physics data was collected over the three-year run. The  
<sup>413</sup> uncertainty on the total integrated luminosity is 1.7% [39].

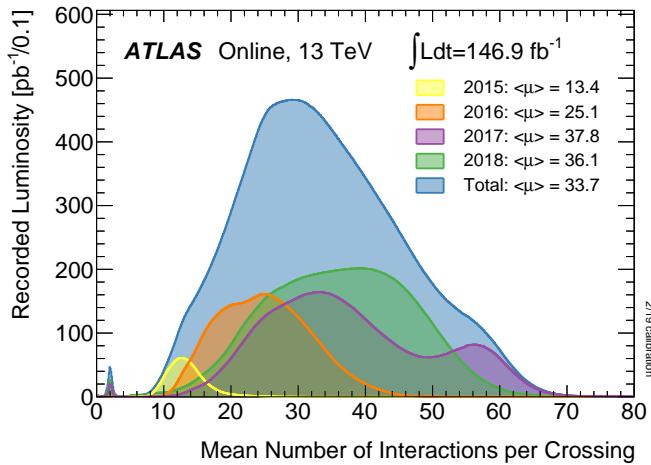
### <sup>414</sup> Pile-up

<sup>415</sup> At the centre of the ATLAS detector, bunches of more than  $10^{11}$  protons meet at a  
<sup>416</sup> small crossing angle. Each bunch-crossing is called an *event*. There is generally at  
<sup>417</sup> most one hard proton-proton scatter per event. Additional interactions are typically  
<sup>418</sup> relatively soft and are known as *pile-up*. Pile-up from interactions within the same  
<sup>419</sup> bunch-crossing is known as *in-time* pile-up while residual signatures from previous  
<sup>420</sup> bunch-crossings is known as *out-of-time* pile-up. The number of pile-up interactions



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

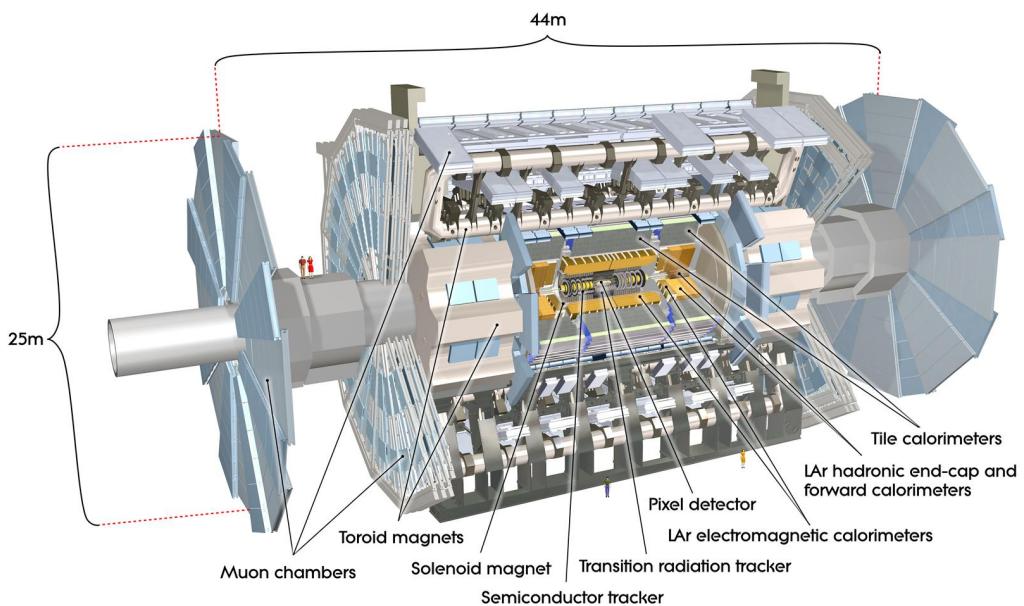
is denoted  $\mu$ , which is often given as a time-averaged value  $\langle \mu \rangle$ . Histograms showing the number of pile-up interactions over the course of Run 2 are shown in Fig. 3.5.



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

### 423 3.3 The ATLAS Detector

424 The ATLAS<sup>2</sup> detector is made up of several specialised sub-detectors which are  
425 arranged concentrically around the nominal interaction point at the centre of the  
426 detector. The detector is designed to cover nearly the entire solid angle around the  
427 collision point. In this section a condensed overview of each sub-detector is given, in  
428 order of increasing radial distance from the point of collision. The inner tracking  
429 detector is described in Section 3.3.1, the electromagnetic and hadronic calorimeters  
430 in Section 3.3.2, the muon spectrometer in Section 3.3.3, and finally the trigger is  
431 described in Section 3.3.4. More complete information on the detector can be found  
432 in Ref. [32], while an overview of physics performance is given in [40].



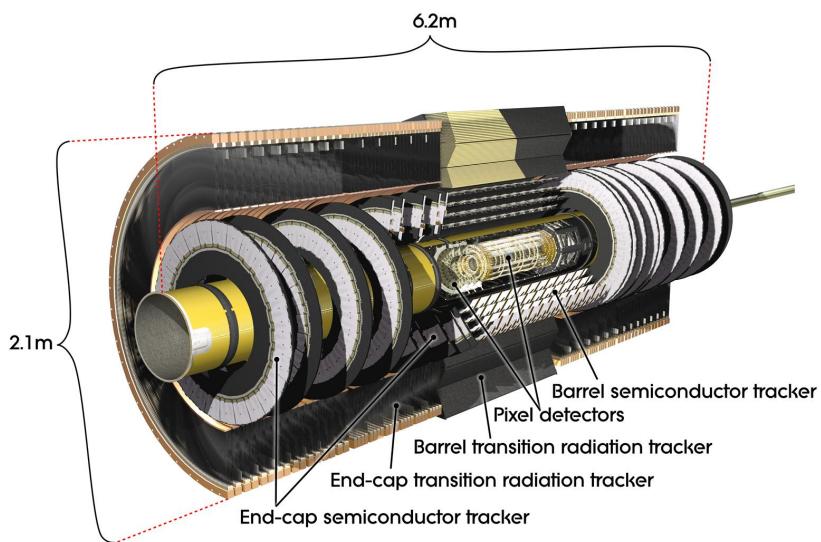
**Figure 3.6:** A 3D model of the entire ATLAS detector [41]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point.

<sup>2</sup>A Toroidal LHC ApparatuS.

### 433 3.3.1 Inner Detector

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

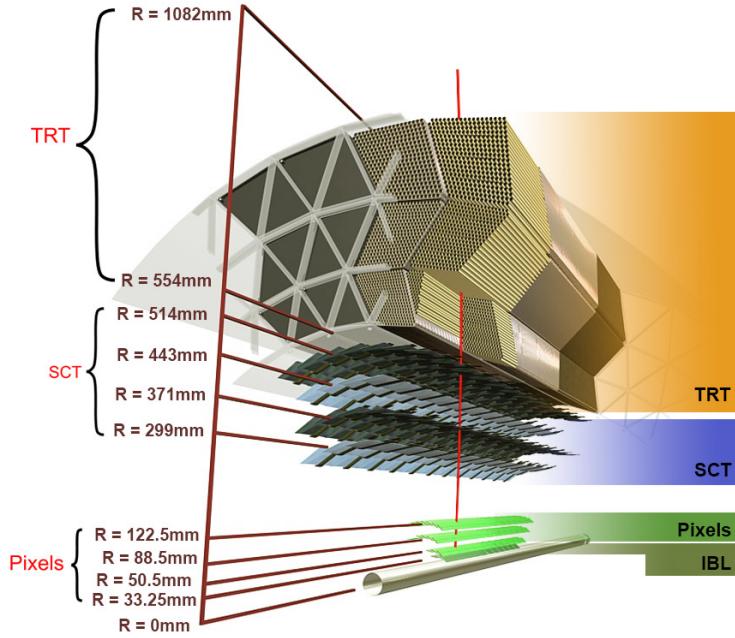
439 The inner detector is made up of several sub-systems, shown in Figs. 3.7 and 3.8. The  
440 high-granularity silicon pixel detector covers the vertex region and typically provides  
441 four spacepoint measurements per track. It is followed by the silicon microstrip  
442 tracker (SCT), which usually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



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

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

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



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

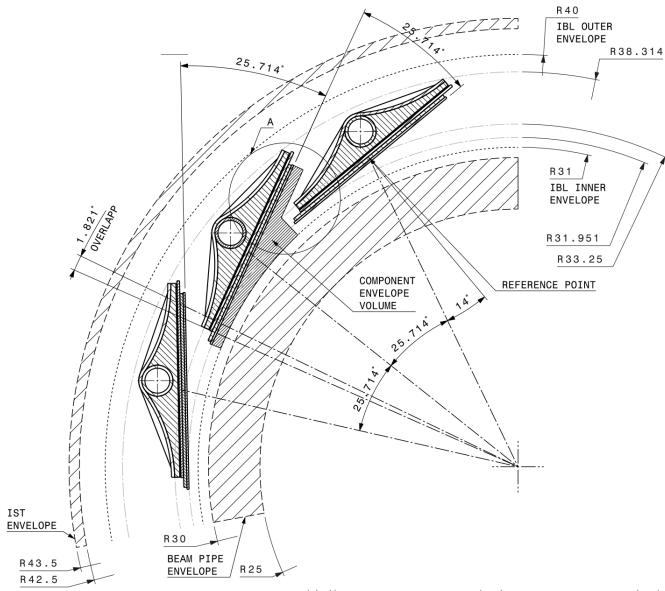
where  $\oplus$  denotes a sum in quadrature. For high- $p_T$  tracks (e.g.  $p_T \approx 100$  GeV) in the central region,  $\sigma(1/p_T) \approx 0.4 \text{ TeV}^{-1}$  corresponding to a relative error of 4%. The momentum resolution generally good enough to correctly identify the sign of the charge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

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

### Pixel Detector

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

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



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

#### 465    Semi-Conductor Tracker (SCT)

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

**476 Transition Radiation Tracker (TRT)**

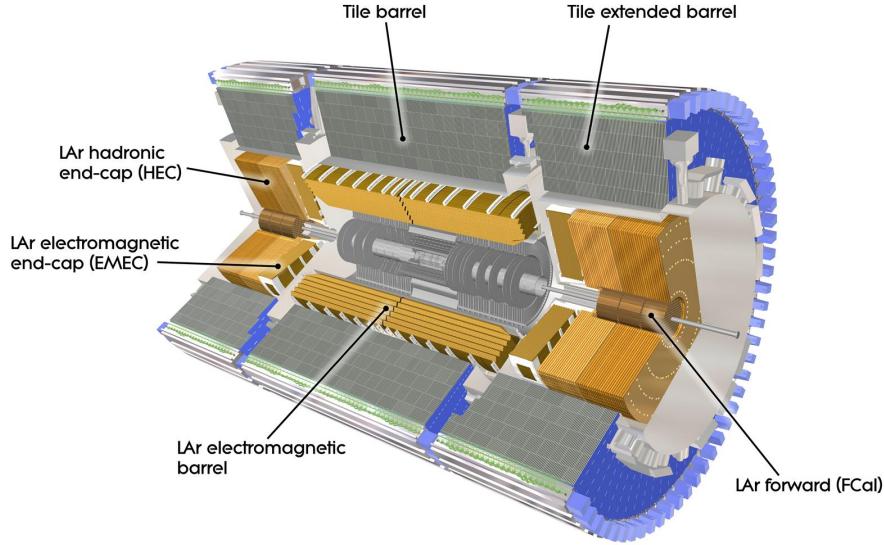
477 The TRT is a straw-tube tracker which complements the higher-resolution silicon-  
478 based tracks by offering a larger number of hits per track (typically around 30) and  
479 a long lever arm, which aids the accurate measurement of particle momentum. It is  
480 made up of approximately 300 000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

**490 3.3.2 Calorimeters**

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

**501 Liquid Argon (LAr) Electromagnetic Calorimeter**

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



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

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

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

### 517 Hadronic Tile Calorimeter

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

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

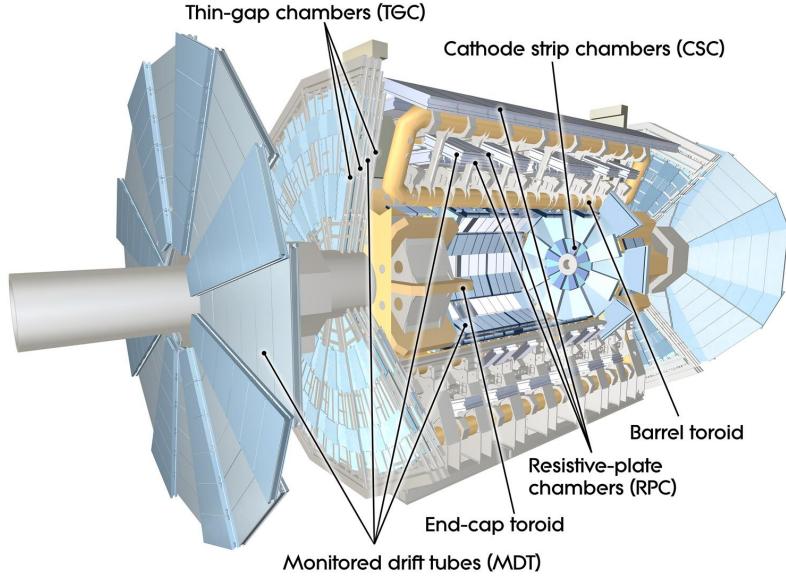
524 The jet energy resolution of the HCal is parameterised as a function of the jet  
 525 transverse energy

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

526 corresponding to a jet energy resolution of 10% at a jet  $p_T$  of approximately 100 GeV  
 527 [49].

### 528 3.3.3 Muon Spectrometer

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



**Figure 3.11:** The ATLAS muon spectrometer [51].

### 546 3.3.4 The Trigger

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

## 562 3.4 Reconstructed Physics Objects

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

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

### 572 3.4.1 Tracks

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

### 580 Space-point Formation (Clustering)

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

**587 Track Finding**

588 Space-points are used to build track seeds. These are groups of three hits which  
589 are geometrically compatible with being part of a track segment. A combinatorial  
590 Kalman filter (KF) is used to build track candidates by extending track seeds. The  
591 filter can create multiple track candidates per seed, with bifurcations along the track  
592 occurring when more than one compatible space-point exists on a given layer. In  
593 this way, the KF creates an excess of *track candidates*, which are only required to  
594 satisfy basic quality requirements. Track candidates are allowed to reuse or *share*  
595 hits freely (a single hit may be used by multiple track candidates). Typically, the  
596 presence of shared hits is a predictor of a bad track due to the high granularity of  
597 the ATLAS tracking detectors. At this stage, there can also be a large number of  
598 incorrect hits assigned to otherwise good tracks, and additionally large number of  
599 *fake* tracks, which are comprised of a majority of wrong hits and do not correspond  
600 to the trajectory of any one physical particle (fake tracks are defined as those where  
601 the majority of associated hits do not originate from one single truth particle, see  
602 Eq. (5.8)). The low quality of tracks at this stage necessitates an ambiguity solving  
603 step, in which candidates are cleaned, and the highest quality track are selected.

**604 Ambiguity Solving**

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

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

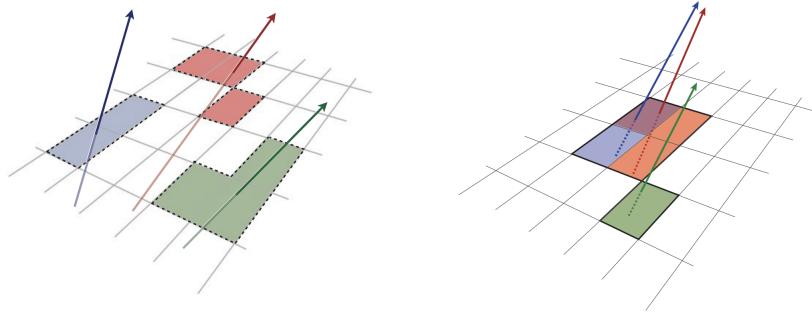
list shared  
hit cut?

## 629 Neural Network Cluster Splitting

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

## 640 Pseudotracking

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



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

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

### 652 3.4.2 Vertices

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

#### 661 Primary Vertices

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

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

667 **Secondary Vertices**

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

679 **3.4.3 Jets**

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

685 Objects can be associated to jets in one of two ways. The first is via a geometrical  
686 matching in  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ , where  $\Delta\eta$  and  $\Delta\phi$  are the differences in  
687 pseudorapidity and azimuthal angle between the jet and the object. The second is  
688 via a ghost association [64], where the object is assigned a negligible momentum and  
689 re-clustered into the jet after its formation.

---

## 690 EMTopo Jets

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

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

694 where  $E_{\text{cell}}$  is the energy measured in a given calorimeter cell, and  $\sigma_{\text{noise, cell}}$  is the  
 695 expected level of noise on the cell (e.g. from pile-up interactions). Topoclusters are  
 696 formed from a seed cell with a large  $S_{\text{cell}}$ , and expanded by iteratively adding neigh-  
 697 bouring cells with a sufficiently large energy significance. Collections of topoclusters  
 698 are then clustered into a jet using the anti- $k_t$  algorithm with a radius parameter of  
 699 0.4 (small- $R$  jets) or 1.0 (large- $R$  jets). More information is available in Ref. [65].

## 700 Particle Flow Jets

701 Particle-flow (PFlow) jets are reconstructed from particle-flow objects [66] using  
 702 the anti- $k_t$  algorithm with a radius parameter of 0.4. Particle-flow objects integrate  
 703 information from both the ID and the calorimeters, improving the energy resolution  
 704 at high transverse momenta and reducing pile-up contamination. The PFlow jet  
 705 energy scale is calibrated according to Ref. [67].

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

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

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

## 717 Large- $R$ Jets

718 Large- $R$  jets have a radius parameter  $R = 1.0$  and are built by clustering topological  
 719 calorimeter clusters using the anti- $k_t$  algorithm [68]. The large radius parameter  
 720 is especially useful for containing the decay products of a boosted Higgs boson, as  
 721 discussed in Chapter 7. Due to their large size, large- $R$  jets benefit from a grooming  
 722 procedure called trimming which remove soft contaminants inside the jet [69, 70].  
 723 Trimming aims to remove jet constituents from pile-up and the underlying event,  
 724 which helps to improve the jet mass resolution and its robustness to varying levels  
 725 of pile-up. The jet mass is computed using a combination of information from the  
 726 calorimeters and ID, and a calibration to data is applied [71].

## 727 Track-jets

728 Track-jets are built by clustering tracks using the anti- $k_t$  clustering algorithm and  
 729 are used in the analysis described in Chapter 7. The radius parameter is allowed  
 730 to vary with transverse momentum such that a broader cone (up to  $R = 0.4$ ) is  
 731 used for low- $p_T$  track-jets and a narrower cone (down to  $R = 0.02$ ) for high- $p_T$   
 732 track-jets [72, 73]. The narrower cone is better suited to clustering highly collimated  
 733 jet constituents at high- $p_T$ . Truth flavour labels for track-jets are derived using the  
 734 same  $\Delta R(\text{hadron}, \text{jet}) < 0.3$  matching scheme as used for PFlow jets.

### 735 3.4.4 Leptons

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

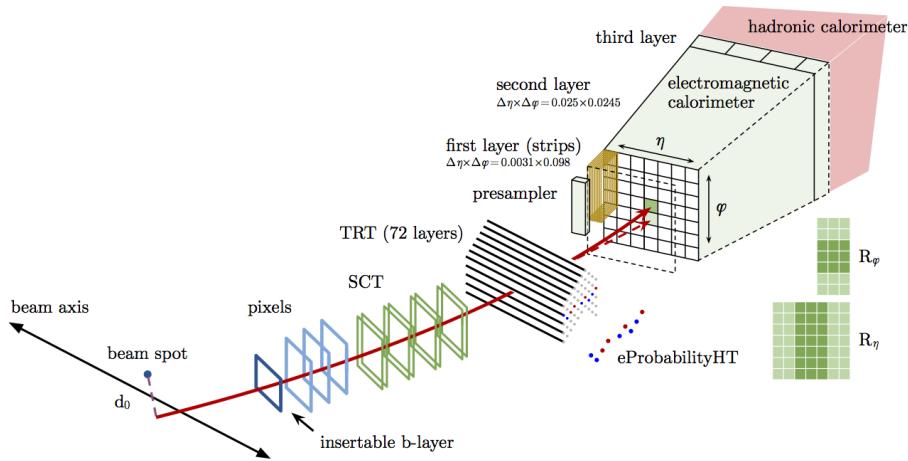
#### 739 Electrons

740 Electrons candidates are reconstructed by matching PV-compatible<sup>3</sup> inner detector  
 741 tracks to topological calorimeter clusters. The track-cluster matching criteria takes  
 742 into account the significant energy loss of the electron due to bremsstrahlung. If a

---

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

743 match is found, a refit of the track is performed using the Gaussian Sum Filter (GSF)  
 744 [74], which better handles trajectory reconstruction in the presence of bremsstrahlung.  
 745 Various identification criteria are then applied to the candidates using a likelihood-  
 746 based (LH) method to improve purity. These include requirements on the track  
 747 quality and cluster matching, the shape of electromagnetic shower in the ECal,  
 748 leakage into the HCal, and the amount of transition radiation detected in the TRT.  
 749 Isolation criteria with respect to other nearby ID tracks and calorimeter clusters may  
 also be applied. A full description can be obtained from Ref. [75].



**Figure 3.13:** A sketch of electron reconstruction using the ATLAS detector [75]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal.

750

## 751 Muons

752 Muon reconstruction makes use of the dedicated MS (see Section 3.3.3), the tracks  
 753 from the ID, and the presence of characteristic signatures in the calorimeters. Muon  
 754 tracks (i.e. a track reconstructed in the MS) are reconstructed by connecting straight-  
 755 line track segments, which are identified via a Hough transform, and combined into  
 756 an approximately parabolic trajectory. Finally, a global  $\chi^2$  fit is performed, taking  
 757 into account possible interactions between the muon and the detector material. A  
 758 reconstructed muon is called *combined* if it completes successful matching to an  
 759 ID track. Combined muons undergo a further fit with the combined ID and MS

760 hits, with the energy loss due to the traversal of the calorimeters being taking into  
761 account.

762 After reconstruction, candidate muons further undergo an identification processes  
763 which helps to efficiently identify prompt muons whilst rejecting background sig-  
764 nals (e.g. non-prompt muons from pion and kaon decays, the punch-through of a  
765 hadron from the calorimeter, or the semi-leptonic decay of a heavy flavour hadron).  
766 Combined muon identification takes into account discrepancies in the  $p_T$  and charge  
767 measurements in the MS and ID, and the  $\chi^2$  of the combined track fit. Selections  
768 on the number of hits in the ID and MS are also applied. At the medium identifi-  
769 cation working point, approximately 96% of muons with  $20 \text{ GeV} < p_T < 100 \text{ GeV}$   
770 are successfully identified. On top of the identification requirements, a number of  
771 isolation requirements can also be applied to further suppress background signals. In  
772 the region  $|\eta| < 2.2$ , the momentum resolution of reconstructed muons is 1.7%.

773 More information on muon reconstruction, identification and isolation can be found  
774 in Ref. [76].

### 775 3.4.5 Missing Transverse Momentum

776 An imbalance in the final state transverse momentum can occur as a result of  
777 incomplete measurement of the final state particles. In particular, neutrinos are  
778 not measured by the detector and contribute to the missing transverse momentum  
779  $\mathbf{E}_T^{\text{miss}}$ . Incomplete detector acceptance and inaccuracies in the reconstruction of the  
780 final state can also contribute to the missing transverse momentum of an event. In  
781 order to calculate the missing transverse momentum, the negative vector sum of  
782 the momentum of all photons, leptons and small- $R$  jets with  $p_T > 20 \text{ GeV}$  is taken.  
783 The momenta of tracks associated to the primary vertex are also taken into account.  
784 The magnitude of  $\mathbf{E}_T^{\text{miss}}$  is written  $E_T^{\text{miss}}$ . More information about missing transverse  
785 momentum reconstruction is provided in [77].

786

# Chapter 4

787

## Tracking and flavour tagging

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

797 The current ATLAS flavour tagger, DL1r [78], is a deep neural network which  
798 accepts as inputs the outputs of a number of independently optimised *low-level*  
799 algorithms [61]. Correspondingly, DL1r is referred to as a *high-level* tagger (i.e. one  
800 that uses a multivariate approach to combine the outputs of the low-level taggers).  
801 Each of these low-level algorithms reconstructs a distinct feature of the experimental  
802 signature of heavy flavour jets using the tracks associated to the jet. The low-level  
803 algorithms are a combination of manually optimised reconstruction algorithms, for  
804 example the SV1 and JetFitter algorithms that reconstruct displaced decay vertices,  
805 and trained taggers such as RNNIP and DIPS that use the IP and hit information  
806 from a variable number of tracks to identify the flavour of the jet [61, 79–81].

807 In addition to DL1r, another widely used high-level tagger is the MV2c10 algorithm  
808 [61, 78, 82]. This tagger is used in the analysis described in Chapter 7. Similar to  
809 DL1r the MV2c10 algorithm takes inputs from the outputs of a number of low-level  
810 algorithms (IPxD, SV1 and JetFitter). The outputs of the low-level algorithms

811 are provided as inputs to a boosted decision tree. The working point is tuned to  
 812 achieve an average  $b$ -jet efficiency of 70% on simulated  $t\bar{t}$  events. At this efficiency  
 813 working point, rejection factors for  $c$ -jets and light-jets are approximately 9 and 304  
 814 respectively.

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

## 820 4.1 $b$ -hadron Reconstruction

821 This section outlines the typical detector signature of a  $b$ -hadron in Section 4.1.1  
 822 and discusses some associated reconstruction difficulties in Section 4.1.2.

### 823 4.1.1 Decay Topology

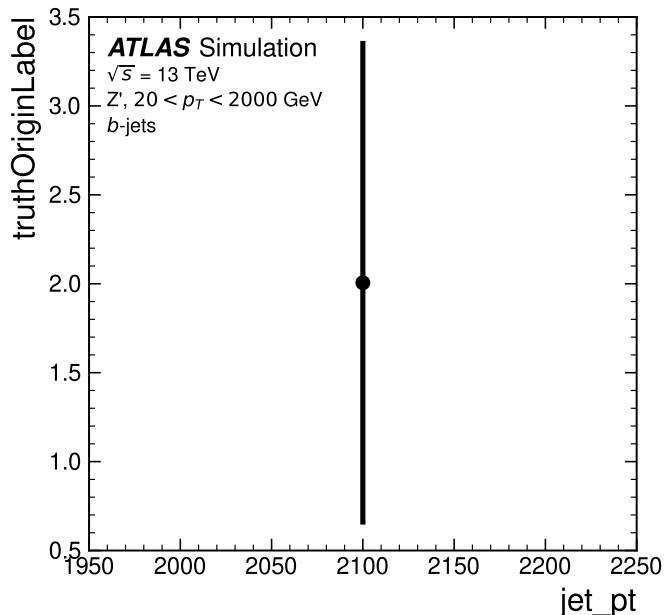
824  $b$ -hadrons are quasi-stable bound states of a bottom quark and one or more lighter  
 825 quarks. Collectively, these are the  $B$ -mesons (e.g.  $B^+ = u\bar{b}$ ,  $B^0 = d\bar{b}$ ) and baryons  
 826 (e.g.  $\Lambda_b^0 = udb$ ). After a  $b$ -quark is produced as the result of a proton-proton collision,  
 827 they quickly hadronise. The hadronisation process is hard – around 70-80% of  
 828 the  $b$ -quark’s momentum is passed to the  $b$ -hadron, with the rest being radiated  
 829 as prompt hadronisation or fragmentation particles. See Ref. [83] for a more in  
 830 depth discussion on hadronisation and the closely related process of fragmentation.  
 831 Henceforth the combined hadronisation and fragmentation products will be referred  
 832 to collectively as fragmentation.

833  $b$ -hadrons are interesting objects of study due to their relatively long proper lifetimes  
 834  $\tau \approx 1.5$  ps [84]. This lifetime corresponds to a proper decay length  $c\tau \approx 450$   $\mu\text{m}$ . In  
 835 the rest frame of the detector, the typical  $b$ -hadron travels a distance

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

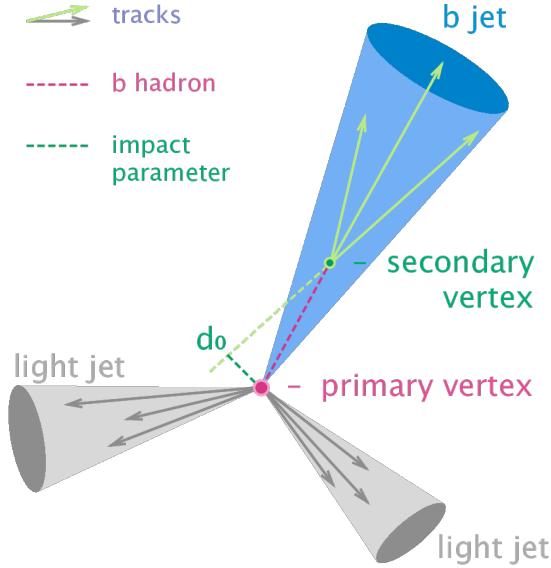
836 before decaying, where in the high energy limit  $\gamma = E_b/m_b$  and  $\beta = v/c = 1$ .

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



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

$b$ -hadrons decay weakly to on average four or five collimated stable particles [85]. These particles, along with any other fragmentation particles, are reconstructed in the detector as a jet. A  $b$ -jet has several characteristic features which differentiate it from light-jets. These features stem from the significant displacement of the  $b$ -hadron that can occurs due to its lifetime. The primary feature is the presence of a high mass secondary vertex that is significantly displaced from the primary vertex. Reconstruction of these vertices from tracks with common points of spatial origin is a common approach used in the identification of  $b$ -jets.



**Figure 4.2:** Diagram of a typical  $b$ -jet (blue) which has been produced in an event alongside two light jets (grey) [86]. 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.

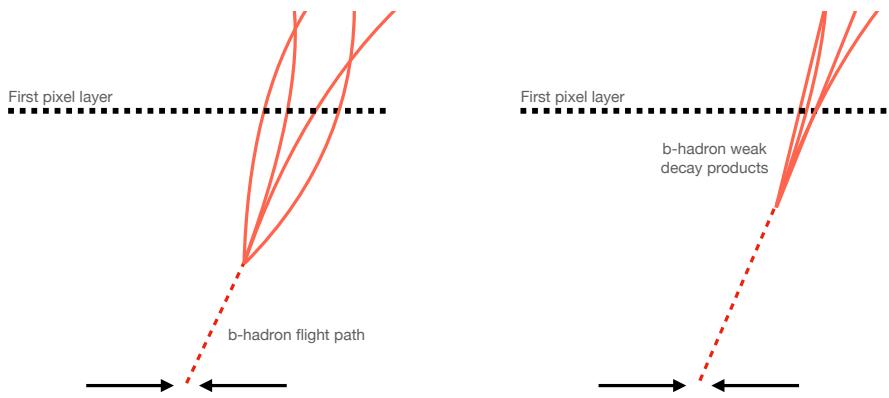
852 Additional signatures of  $b$ -hadrons are as follows. Associated tracks and SVs can have  
 853 a large transverse impact parameter  $d_0$  as a result of the  $b$ -hadron displacement (as  
 854 shown in Fig. 4.2). Since it is common for the  $b$ -hadron to decay to a  $c$ -hadron with  
 855 non-negligible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

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

#### 863 4.1.2 Challenges Facing $b$ -hadron Reconstruction

864 As discussed, a necessary requirement for successful  $b$ -tagging is the efficient and  
 865 accurate reconstruction of the charged particle trajectories in the jet. For high  $p_T$  jets

( $p_T > 200$  GeV) this task becomes difficult due to a combination of effects. As the  $b$ -hadron energy increases, the multiplicity of the fragmentation products inside the jet increases, while the multiplicity of the products of the weak decay is unaffected. The “signal” tracks (those from the weak decay of the  $b$ -hadron) therefore become outnumbered. Both fragmentation and  $b$ -hadron weak decay products also become increasingly collimated as their inherited transverse momentum increases. At high energies, the increased decay length of  $b$ -hadrons (and  $c$ -hadrons) means that decay products have less of an opportunity to diverge before reaching the first tracking layers of the detector (shown in Fig. 4.3). If the weak decay of the  $b$ -hadron takes place close enough to a detector layer, or if the particles are otherwise sufficiently collimated, charge deposits left by nearby particles may not be resolved individually, instead being reconstructed as merged clusters.

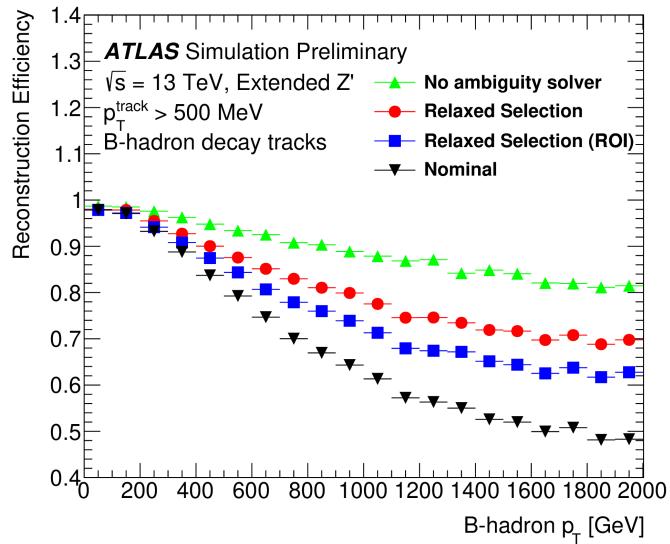


**Figure 4.3:** At lower  $p_T$  (left) the decay length of the  $b$ -hadron is reduced, and the resulting decay tracks are less collimated. At higher  $p_T$  (right) the  $b$ -hadron decay length increases and the resulting decay tracks are more collimated and have less distance over which to diverge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

As discussed in Section 3.4.1, merged clusters are generally rare, and so shared hits generally predict bad tracks and 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 to be effectively dealt with but in the standard reconstruction the presence of these can end up impairing the successful reconstruction of the track. Furthermore, decays may also take place inside the tracking detectors themselves, which at best leads to missing

886 measurements on the most sensitive detector layers, and at worst can lead to wrong  
 887 inner layer hits being added to displaced tracks, since the reconstruction process  
 888 penalises tracks without inner layer hits.

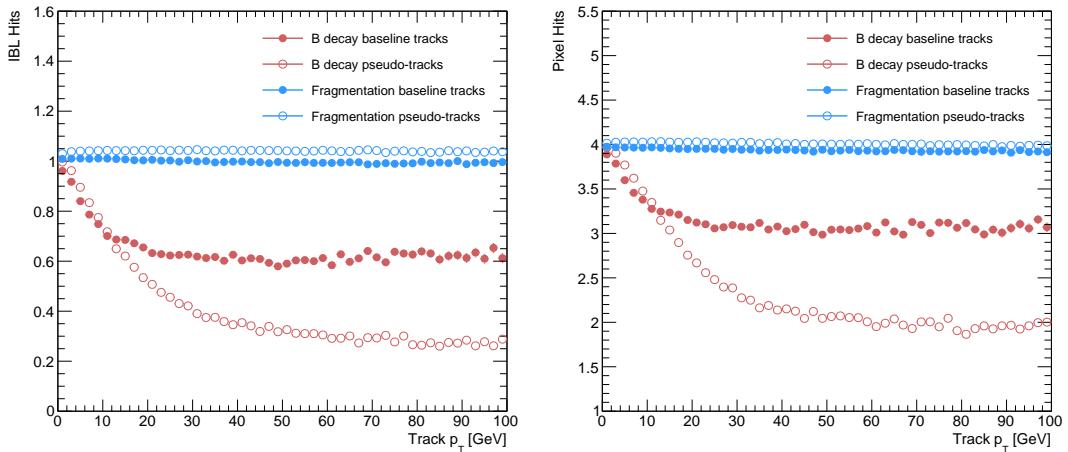
889 The above effects create two related, but distinct problems for  $b$ -tagging. The first  
 890 part is a drop in track reconstruction efficiency. The presence of shared and missing  
 891 hits reduces a track's score in the ambiguity solver meaning that higher ranking, but  
 892 potentially worse, track candidates are processed first and take ownership of the hits.  
 893 This can make it difficult for otherwise reasonable  $b$ -hadron decay tracks to meet  
 894 the ambiguity solver's stringent track quality requirements, leading to their rejection  
 895 at this stage and an overall decrease in the  $b$ -hadron decay track reconstruction  
 896 efficiency as shown in Fig. 4.4.



**Figure 4.4:**  $b$ -hadron decay track reconstruction efficiency as a function of truth  $b$ -hadron  $p_T$  [87]. Nominal track reconstruction is shown in black, while the track reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver efficiency) is shown in green. For high- $p_T$   $b$ -hadrons, the ambiguity solver is overly aggressive in its removal of  $b$ -hadron decay tracks. Suggestions for the improvement of the track reconstruction efficiency in this regime by the loosening of cuts in the ambiguity solver are shown in blue and red.

897 The second part of the problem is that, due to the high multiplicity of clusters available  
 898 for assignment in the vicinity of the typical high energy  $b$ -hadron decay track, and  
 899 also given the strong positive bias of the ambiguity solver towards those tracks  
 900 with pixel measurements in each layer (especially the innermost IBL measurement),

many  $b$ -hadron decay tracks are assigned incorrect inner layer hits. This is only a problem for those decay products which were produced within the pixel detector as a result of a significantly displaced  $b$ -hadron decay, and so do not have a correct hit available for assignment. Fig. 4.5 shows the number of hits as a function of the reconstructed track  $p_T$  for fragmentation tracks and tracks from the weak decay of the  $b$ -hadron. The baseline tracks represent the standard reconstruction setup, while the pseudotracks represent the ideal tracking setup as outlined in Section 3.4.1. The incorrect hits may skew the parameters of the track, which can in turn mislead the downstream  $b$ -tagging algorithms. In particular,  $b$ -tagging algorithms rely heavily on the transverse impact parameter significance  $s(d_0)$  of the track. The quality of this measurement is expected to be adversely affected by wrong inner-layer hits on the track. Furthermore, multiple tracks sharing an incorrect hit can lead to the creation of spurious secondary vertices, which can cause further problems for the downstream  $b$ -tagging algorithms.



**Figure 4.5:** Hit multiplicities on the IBL (left) and the pixel layers (right) as a function of the  $p_T$  of the reconstructed track. Tracks from the weak decay of the  $b$ -hadron are shown in red, while fragmentation tracks (which are prompt) are in blue. Baseline tracks are those produced in the standard reconstruction described in Section 3.4.1, while pseudotracks represent the ideal performance of the ATLAS detector and are described in Section 3.4.1. Hit multiplicities on the pseudotracks decrease at high  $p_T$  due to the flight of the  $b$ -hadron before its decay. The baseline tracks have more hits than the pseudotracks, indicating that they are being incorrectly assigned additional hits on the inner layers of the detector.

The combination of the effects described makes reconstructing tracks in the core of high  $p_T$   $b$ -jets particularly challenging. The reduced reconstruction efficiency of  $b$ -hadron decay tracks and incorrectly assigned hits is thought to be the primary

918 cause of the observed drop in  $b$ -tagging efficiency at high energies, however further  
919 study is required to determine which effect may dominate.

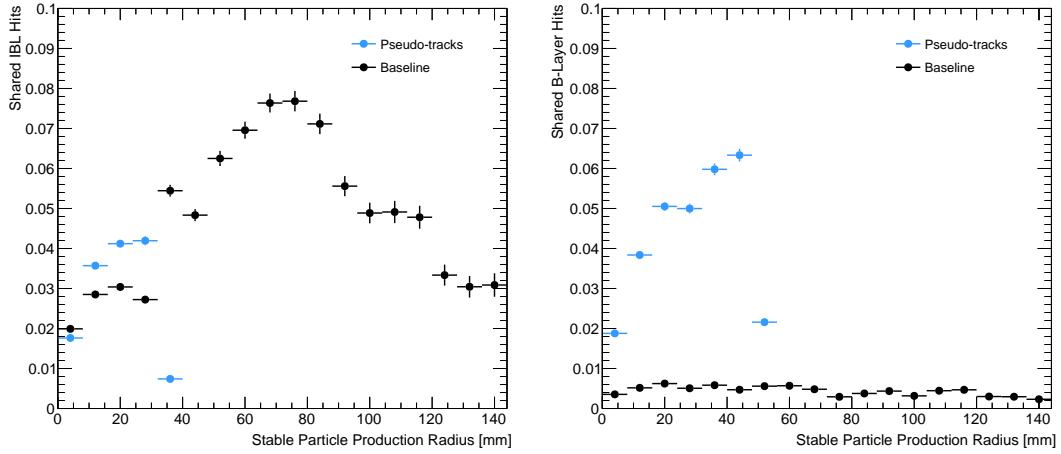
include plot from sebs study showing they are approx similar impacts? or just mention result? Can do put need to remove ATLAS labels. Alternatively you can put an internal reference to his work and state what the outcome is

## 920 4.2 Investigations into High $p_T$ $b$ -hadron Tracking

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

### 926 4.2.1 Shared Hits

927 The ambiguity solver is not run for pseudotracks. However, if the standard track  
928 collection is produced alongside the pseudotracks, then cluster splitting neural  
929 networks will be run for the standard tracks, and the resulting classification of  
930 clusters will be propagated to hits on pseudotracks. This quirk allows one to study  
931 the inefficiencies of the cluster splitting process, and relatedly to determine whether  
932 shared hit cuts in the ambiguity solver are too loose or too tight. The fraction of hits  
933 that are shared for the IBL and the B-layer is shown in Fig. 4.6. The shared hits on  
934 pseudotracks represent correctly assigned hits from merged clusters that were not  
935 able to be classified as split by the cluster splitting neural networks. As such, these  
936 represent the number of shared hits the ambiguity solver should aim to allow given  
937 the current performance of the cluster splitting algorithm. For shared hits on the IBL  
938 for particles produced before the IBL, the baseline selection appears to be successful  
939 in disallowing excessive numbers of shared hits. However, the ambiguity solver fails to  
940 limit shared hits for those particles produced after the IBL, reflecting the previously  
941 discussed problem of displaced tracks picking up incorrect hits. Meanwhile, it is clear  
942 that for the B-layer, the ambiguity solver is being overly aggressive in its rejection of  
943 shared hits.



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

#### 944 4.2.2 Global $\chi^2$ Fitter Outlier Removal

945 This section documents ongoing progress into improvement of hit-to-track assignment  
 946 by using the Global  $\chi^2$  Fitter (GX2F) to identify and prevent incorrect hits from  
 947 being assigned to tracks during the track fit. This is in contrast to a previously  
 948 investigated approach [88] which attempted to identify and remove wrong hits after  
 949 the reconstruction of the track. As part of the track fit, an outlier removal procedure  
 950 is run, in which suspicious hits are identified and removed.

951 The GX2F code, as a relatively low-level component of track reconstruction, has  
 952 not undergone significant modification for several years, and was originally only  
 953 optimised in the context of prompt, isolated tracks. During this time, a new tracking  
 954 sub-detector, the IBL, was installed. The motivation for looking at the GX2F is that  
 955 these changes may require re-optimisation of the GX2F code, and in particular the  
 956 outlier removal procedures. Further motivation for this approach comes from the low  
 957 rate of labelled outliers in baseline tracking. For example, while approximately 15%  
 958 of  $b$ -hadron decay tracks have a wrong IBL hit (a value which only increases with  
 959 the  $p_T$  of the  $b$ -hadron), less than 1% of this tracks have had their IBL hit labelled  
 960 and removed as an outlier.

---

**961 Implementation**

**962** The outlier removal procedure for the pixel detector is described in this section.  
**963** The hits on the track are looped over in order of increasing radial distance to the  
**964** beam pipe. For each hit, errors  $\sigma(m_i)$  on the measurement of the transverse and  
**965** longitudinal coordinates are calculated. These errors are dependent on the sub-  
**966** detector which recorded the measurement (some sub-detectors are more precise than  
**967** others). Additionally, a residual displacement  $r_i = m_i - x_i$  between the predicted  
**968** position of the track  $x_i$  (inclusive of the current measurement), and the position of  
**969** the hit itself,  $m_i$ , is calculated. The pull  $p_i$  on the track state due to the current  
**970** measurement is calculated according to

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

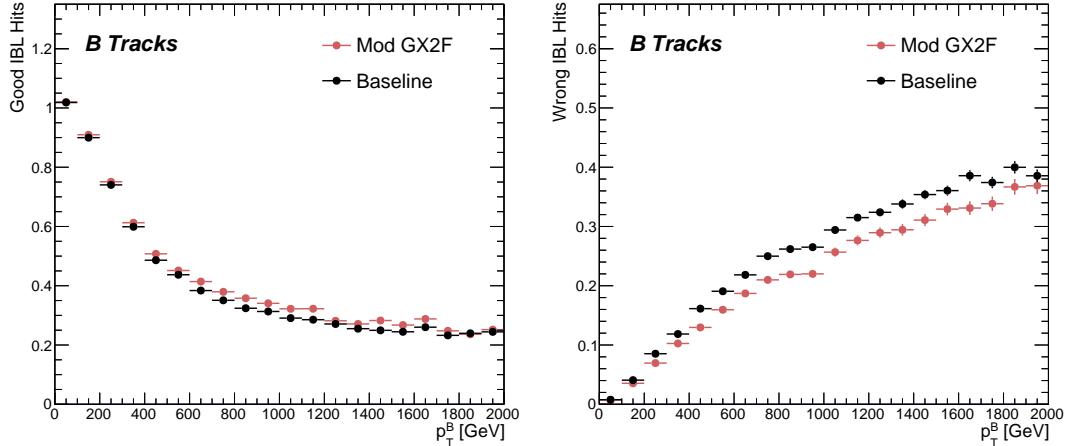
**971** This pull is computed for the transverse and longitudinal coordinates of the mea-  
**972** surement, and the maximum of the two is selected and checked to see if it exceeds  
**973** a certain selection threshold. If it does, the hit will be removed if the track also  
**974** exceeds a threshold on the total  $\chi^2/n$ . The results of varying the outlier selection  
**975** and  $\chi^2/n$  thresholds are described below.

**976 Cut Optimisation**

**977** A systematic variation of the outlier selection and  $\chi^2/n$  thresholds has been carried  
**978** out. Both thresholds were reduced in fixed step sizes of 0.25 for the outlier selection  
**979** threshold and 1 for the  $\chi^2/n$  threshold. The results for the best performing selections  
**980** are discussed below. The value of the outlier selection threshold was reduced from 4  
**981** down to 1.75, a change which affects all silicon layers (the TRT has separate outlier  
**982** removal logic). Furthermore, a specific cut for the IBL was introduced, and is set  
**983** to 1.25. The second threshold on the track  $\chi^2/n$  was also reduced from 7 to 4.  
**984** Finally, instead of taking the maximum of the pulls in the longitudinal and transverse  
**985** directions, a quadrature sum is taken of these two values and used. This variation is  
**986** labelled “Mod GX2F” in plots.

**987** The results are shown in Fig. 4.7 and demonstrate a reduction in wrong hit assignment  
**988** whilst also improving slightly the rate at which good hits are assigned to tracks. For a

989 1 TeV track, the rate to assignment good hits to the track increases by approximately  
 990 10%, while the rate to assign incorrect hits decreases by approximately 16%. The  
 991 improvements are also observed when looking inclusively in all tracks, which avoids  
 992 the need for a specific  $b$ -jet region-of-interest selection.

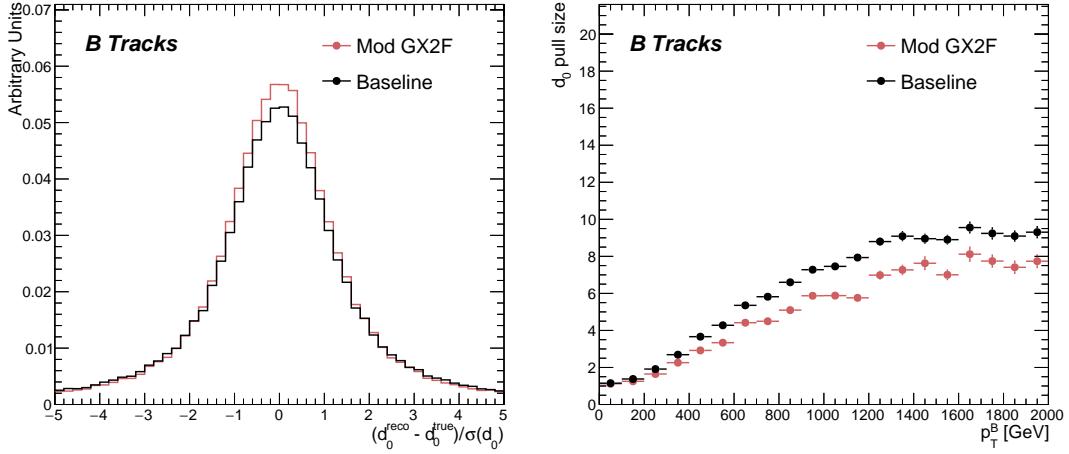


**Figure 4.7:** The rate to assign good (left) and wrong (right) IBL hits as a function of  $b$ -hadron  $p_T$  for tracks using baseline tracking (black) and the modified version of the outlier removal procedure (red). For each track, the corresponding  $p_T$  bin is filled with the number of good or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

993 An improvement, though modest, of all track parameter resolutions and pulls is  
 994 observed. The improvement for the transverse impact parameter pull is shown in  
 995 Fig. 4.8. The results demonstrate an improvement in hit assignment, unchanged  
 996 reconstruction efficiency, and modest improvement in track parameter resolutions  
 997 and pulls. In addition, the truth match probability of track is unchanged, suggesting  
 998 that there is no increase in fake track rates. The changes are expected to have a  
 999 negligible impact on computational resources.

### 1000 4.3 Conclusion

1001 In this section, the difficulties facing efficient and accurate track reconstruction,  
 1002 and hence performant  $b$ -tagging, have been outlined. The ambiguity solver, which  
 1003 attempts to clean or reject tracks which have an excessive number of shared hits,  
 1004 is shown to be overly aggressive in the removal of  $b$ -hadron decay product track  
 1005 candidates. The ambiguity solving process relies on a complicated pre-defined



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

selection which has not been optimised for high transverse momentum  $b$ -hadron track reconstruction. These conclusions have motivated further ongoing studies into the improvement of the track reconstruction in dense environments and the high- $p_T$  regime, such as those in Ref. [87].

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

1021 **Chapter 5**

1022 **Track Classification MVA**

1023 The chapter details work on implementing a multivariate algorithm (MVA) to predict  
1024 the truth origin of reconstructed tracks. An introduction to formalisms of machine  
1025 learning is given in Section 5.1. In Section 5.3, the truth origin label is defined,  
1026 and in Section 5.4 these labels are used to train a machine learning model that can  
1027 effectively discriminate between good and fake tracks. Several studies motivated this  
1028 work by demonstrating that at high  $p_T$ ,  $b$ -tagging performance was degraded by the  
1029 presence of large numbers of poorly reconstructed or fake tracks. If an algorithm  
1030 could be trained to detect fake tracks, these could be removed before their input to  
1031 the  $b$ -tagging algorithms with the aim of improving performance.

1032 **5.1 Machine Learning Background**

1033 Over the past few decades, machine learning (ML) techniques have become increasing-  
1034 ly popular in high energy physics experiments due the increased volumes of  
1035 high-dimensional data and improvements in the techniques used (in particular deep  
1036 learning). Machine learning is the process by which a computer program uses data  
1037 to learn suitable parameters for a predictive model. This is opposed to explicitly  
1038 providing instructions on how to perform a task. A subfield known as *supervised*  
1039 *learning* is used in this work, and consists of exposing a model to a large number of  
1040 labelled examples in order to extract relationships between the input data and their  
1041 labels. These relationships are often complex, and explicitly programmed rules can  
1042 fail to fully capture the relationships between inputs and outputs.

In the simplest case, a set of  $m$  labelled training examples  $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$  is collected. Each element  $(x_i, y_i)$  consists of a input vector  $x_i \in \mathbb{R}^{\text{input}}$ , and the corresponding label  $y_i$ . In classification problems, these labels are integer *class labels*  $y_i \in \{0, \dots, N - 1\}$ , where  $N$  is the number of classes, which specify which of a pre-determined set of categorical classes the training example belongs to. The rest of the discussion in this chapter is limited to binary classification problems ( $N = 2$ ). The two classes are often referred to as signal ( $y_i = 1$ ) and background ( $y_i = 0$ ), which need to be separated. Collecting sufficient and suitable data is one of the primary challenges of machine learning, as such data is not always readily available. Fortunately, sophisticated tools to simulate particle collisions have already been developed by the scientific community [89, 90]. These tools play a key role in generating a suitably large amount of labelled data which is used to train algorithms. More detail on the input datasets is given in Section 5.4.1.

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

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

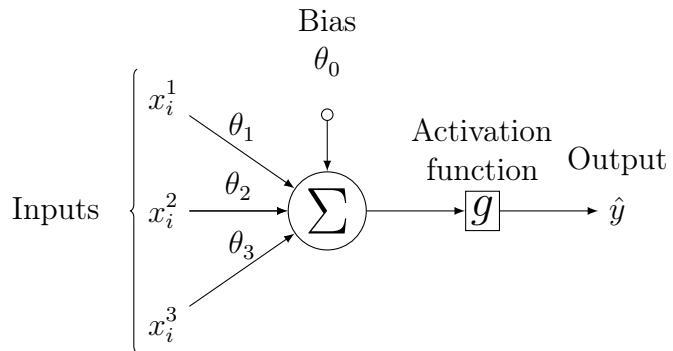
The output  $\hat{y}_i$  is in the interval  $(0, 1)$  so as to be interpreted as the probability that the input example  $x_i$  belongs to the signal class. The parameters  $\theta$  of the model are randomly initialised, and the model is designed to be expressive enough to correctly map the inputs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

### 5.1.1 Neural Networks

Neural networks (NNs) are a common choice for the machine learning model  $f$  since they have the ability to approximate any function [91] and are easy to train via backpropagation [92].

1072 **Artificial Neurons**

1073 The basic functional component of a NN is the *artificial neuron* or node, which is  
1074 loosely inspired by a mathematical model of a biological neuron [93, 94]. A diagram  
1075 of an artificial neuron is shown in Fig. 5.1 Each neuron is defined by its parameters  
1076 or *weights*  $\theta$  and a choice of activation function. Each neuron takes a fixed number  
1077 of inputs and computes the dot product of the input and weight vectors  $x^T \theta$  and  
1078 additionally adds a constant bias term  $\theta_0$ . This term plays the role of a trainable  
1079 constant value that is independent of the inputs.

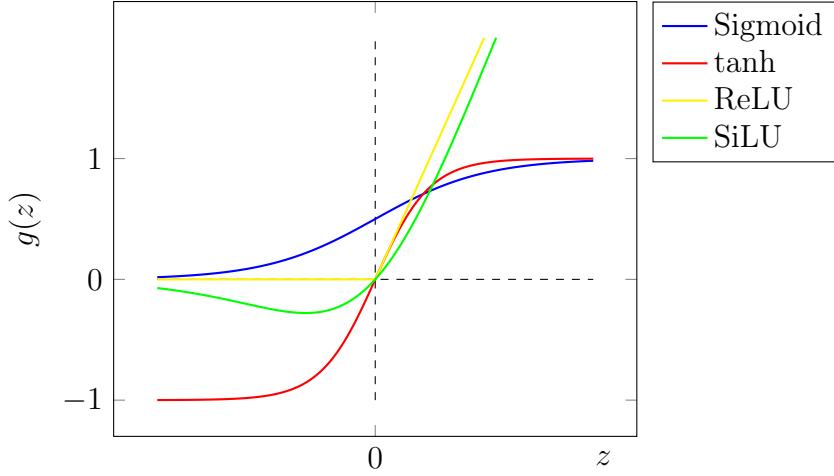


**Figure 5.1:** A diagram displaying the logical flow of a single neuron with three inputs  $x_i^j$ . Each input is multiplied by a weight  $\theta_j$ , and the resulting values are summed. A bias term  $\theta_0$  is added, and the result  $z$  is passed to an activation function. Each neuron can be thought of as a logistic regression model.

1080 The output of the dot product and bias term  $z$  is fed into an activation function  
1081  $g$ . The activation function has several uses, most notably acting as a source of  
1082 non-linearity and bounding the output of the neuron. Some common activation  
1083 functions (sigmod, tanh, ReLU and SiLU) are shown in Fig. 5.2. The choice of  
1084 activation function can have implications for the performance and convergence of  
1085 the network, since the gradient of  $g$  is used to compute the weight updates during  
1086 training. This is also why input data is typically normalised to have zero mean and  
1087 unity variance [95].

1088 **Networks**

1089 Several neurons are linked together in layers to form a neural network. The inputs  
1090 are propagated layer-by-layer through the network until reaching the final output



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

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

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

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

### 5.1.2 Training with Gradient Descent

A training algorithm is used to optimise the weights of a NN after exposure to the training data. The training algorithm works by minimising a loss function  $L$ , which quantifies the error in the model's predictions. NNs are commonly trained using backpropagation in combination with a variant of the stochastic gradient descent algorithm to iteratively update the model parameters. In binary classification

1106 problems, the binary cross entropy given in Eq. (5.3) loss if often used.

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

1107 Since the model  $f$  is differentiable, the error for each parameter  $\theta_i$  can be computed by  
1108 taking the partial derivative of  $L$  with respect to the parameter. Updated parameters  
1109  $\theta'_i$  are calculated by updating the original parameter in the direction which reduces  
1110 the loss.

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

1111 The hyperparameter  $\alpha$  is known as the *learning rate* and dictates the size of the  
1112 step taken in the direction of the slope. The errors for each parameter are efficiently  
1113 calculated using the backpropagation algorithm [92]. The process of updating weights  
1114 is repeated until the weights converge, which means the network is trained. In practice,  
1115 small batches of the input data are shown to the network at a time. For each batch  
1116 the average loss is calculated and the network's weights are updated. There are many  
1117 extensions and variations of the gradient descent algorithm. This work uses the Adam  
1118 optimiser which adds momentum to the weight updates (dampening oscillations)  
1119 and an adaptive per-parameter learning rate [96].

## 1120 5.2 Graph Neural Network Theory

1121 Graph neural networks are a more sophisticated neural network model (see Sec-  
1122 tion 5.1.1) that are designed to operate on graph structured data. A brief introduction  
1123 to GNNs is provided in this section following the formalism in Ref. [97].

1124 A graph  $\mathcal{G}$  consists of a set of  $N^n$  nodes  $\mathcal{N} = \{h_i\}_{i=1:N^n}$ , a set of  $N^e$  edges  $\mathcal{E} =$   
1125  $\{e_i\}_{i=1:N^e}$ , and a global representation  $u$ . Each node represents an individual object,  
1126 and edges are directed connections between two nodes, called the *sender* and *receiver*  
1127 nodes. The connectivity of the graph therefore encodes information about the  
1128 relationships between objects that exist in the graph.

1129 A single graph network layer consists of three separate update functions  $\phi^e$ ,  $\phi^h$  and  
1130  $\phi^u$  one for each of the nodes, edges, and global graph representation, and similarly  
1131 three aggregation functions  $\rho^{e \rightarrow h}$ ,  $\rho^{e \rightarrow u}$  and  $\rho^{h \rightarrow u}$ . The aggregation functions combine

information across different edges or nodes for their input into the update functions, which produce new representations for the nodes, edges and global objects based on the information in the previous layer and the aggregated information. The update functions are typically each implemented as a dense feedforward neural network (as described in Section 5.1.1). The edges  $e_i$  are updated by a edge network  $\phi^e$  as in

$$e'_i = \phi^e(e_i, h_s, h_r, u), \quad (5.5)$$

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

$$h'_i = \phi^h(\bar{e}'_i, h_i, u), \quad (5.6)$$

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

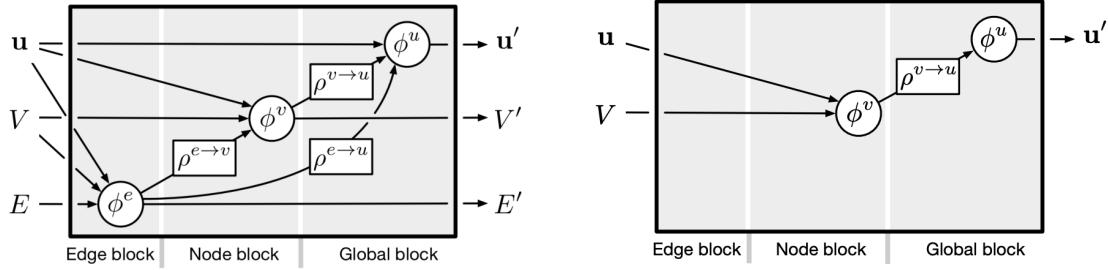
$$u' = \phi^u(\bar{e}', \bar{h}', u), \quad (5.7)$$

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

The graph network layer performs a graph convolution, in an analogous way to a convolutional neural network operating on a grid of pixels. The above description is general, and not all concrete implementations of GNNs need implement every aspect. For example, the global graph representation need not be present, and it is also possible that no dedicated edge features are present. In such cases the corresponding update and aggregation functions are not needed. Fig. 5.3 shows two possible graph network update layers. The layer used in the GN1 model is specified in more detail in Section 6.3.3.

### 5.3 Track Truth Origin Labelling

Crucial to supervised learning techniques are the ground truth class labels which the machine learning model is trained to predict. A set of track truth labels which a



**Figure 5.3:** The flow of information through a graph neural network layer for (left) a full implementation of the layer and (right) a deep sets model [98]. Reproduced from Ref. [97].

high degree of granularity have been implemented in the ATLAS software stack, and are listed in Table 5.1. The labelling scheme has been designed to be useful beyond the classification of good and fake tracks. The origins are determined by analysing the simulated record to determine the physical process that led to the creation of the truth (i.e. simulated) particle which is associated with each reconstructed track. Tracks are associated with truth particles by selecting the particle with the highest *truth-matching probability* (TMP), defined in Eq. (5.8). This is a weighted sum of the number of hits on a reconstructed track which are from the same truth particle, versus the total number of hits on the track. The weights are subdetector-dependent and are designed to account for the varying importance of the different subdetectors (based upon their precision) in the reconstruction of a track.

$$\text{TMP} = \frac{10N_{\text{Pix}}^{\text{good}} + 5N_{\text{SCT}}^{\text{good}} + N_{\text{TRT}}^{\text{good}}}{10N_{\text{Pix}}^{\text{all}} + 5N_{\text{SCT}}^{\text{all}} + N_{\text{TRT}}^{\text{all}}} \quad (5.8)$$

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

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

| Truth Origin   | Description  |
|----------------|--|
| Pileup         | From a $pp$ collision other than the primary interaction                     |
| Fake           | Created from the hits of multiple particles                                  |
| Primary        | Does not originate from any secondary decay                                  |
| fromB          | From the decay of a $b$ -hadron  |
| fromBC         | From a $c$ -hadron decay which itself is from the decay of a $b$ -hadron     |
| fromC          | From the decay of a $c$ -hadron which is not from the decay of a $b$ -hadron |
| OtherSecondary | From other secondary interactions and decays                                 |

**Table 5.1:** Truth origins which are used to categorise the physics process that led to the production of a track. Tracks are matched to charged particles using the truth-matching probability [56]. A truth-matching probability of less than 0.5 indicates that reconstructed track parameters are likely to be mismeasured and may not correspond to the trajectory of a single charged particle. The “OtherSecondary” origin includes tracks from photon conversions,  $K_S^0$  and  $\Lambda^0$  decays, and hadronic interactions.

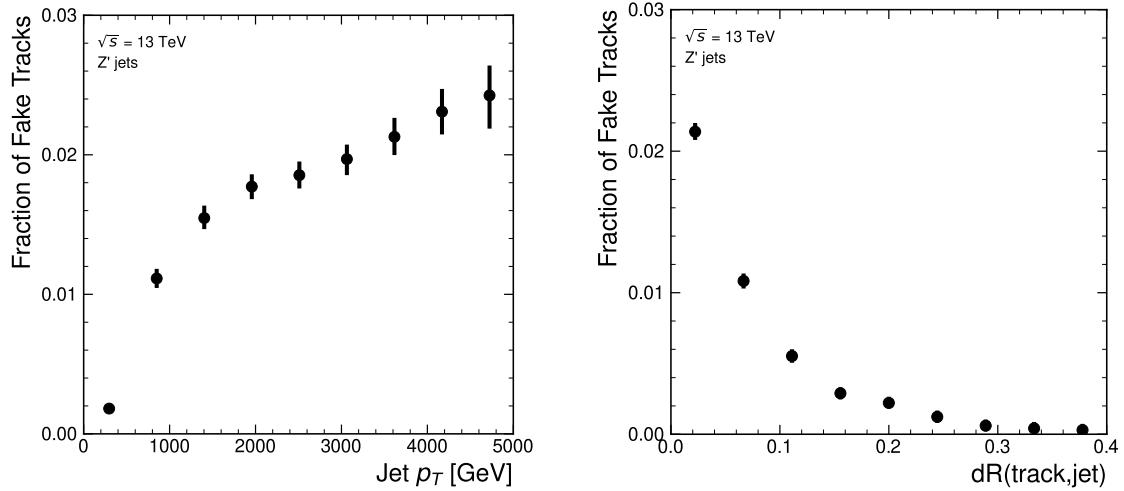
## 5.4 Fake Track Identification Tool

The rate of fake tracks increases at high transverse momentum as shown in Fig. 5.4 due to the difficulties in track reconstruction outlined in Section 4.1.2. The performance of  $b$ -tagging algorithms is reduced as a direct result of the presence of these tracks as shown for SV1 (see Section 3.4.2) in Fig. 5.5, where the light-jet efficiency decreases by up to 35% at a  $b$ -jet efficiency of 35%.

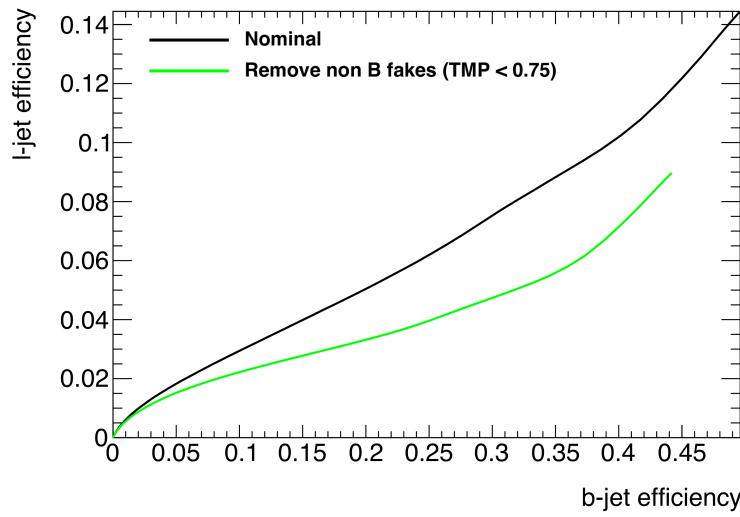
To identify and remove fake tracks, a NN classification tool was trained with all non-fake tracks as the signal class and fake tracks as the background class. Inputs to the model are described in Section 5.4.2, while fake track removal performance is given in Section 5.4.4.

### 5.4.1 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 democratically to equal numbers of  $b$ -,  $c$ - and light-jets. The generation of the simulated event samples includes the effect of multiple  $pp$



**Figure 5.4:** Rate of fake tracks as a function of jet transverse momentum (left) and  $\Delta R(\text{track},\text{jet})$  (right). The rate of fake tracks increases significantly as a function of  $p_T$ , and also increases as the distance to the jet axis decreases.



**Figure 5.5:** The light-jet efficiency of the low level tagger SV1 for jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV, as a function of  $b$ -jet efficiency. The nominal tracking setup (black) is shown alongside the case where fake tracks which are not from the decay of a  $b$ -hadron are removed. The light-jet efficiency is decreased, demonstrating that the presence of fake tracks is detrimental to the algorithm performance.

1191 interactions per bunch crossing with an average pileup of  $\langle \mu \rangle = 40$ , which includes  
1192 the effect on the detector response due to interactions from bunch crossings before  
1193 or after the one containing the hard interaction.

1194 The  $t\bar{t}$  events are generated using the POWHEGBox [99–102] v2 generator at next-  
1195 to-leading order with the NNPDF3.0NLO [103] set of parton distribution func-  
1196 tions (PDFs). The  $h_{\text{damp}}$  parameter<sup>2</sup> is set to 1.5 times the mass of the top-quark  
1197 ( $m_{\text{top}}$ ) [104], with  $m_{\text{top}} = 172.5$  GeV. The events are interfaced to PYTHIA 8.230 [105]  
1198 to model the parton shower, hadronisation, and underlying event, with parameters  
1199 set according to the A14 tune [106] and using the NNPDF2.3LO set of PDFs [107].  
1200  $Z'$  events are generated with PYTHIA 8.2.12 with the same tune and PDF set. The  
1201 decays of  $b$ - and  $c$ -hadrons are performed by EVTGEN v1.6.0 [108]. Particles are  
1202 passed through the ATLAS detector simulation [109] based on GEANT4 [110].

1203 Jets are required to have a pseudorapidity  $|\eta| < 2.5$  and  $p_{\text{T}} > 20$  GeV. Additionally, a  
1204 standard selection using the Jet Vertex Tagger (JVT) algorithm at the tight working  
1205 point is applied to jets with  $p_{\text{T}} < 60$  GeV and  $|\eta| < 2.4$  in order to suppress pile-up  
1206 contamination [111].

### 1207 5.4.2 Model Inputs

1208 The fake track MVA is given two jet variables and 20 tracking related variables  
1209 for each track fed into the network. The jet transverse momentum and signed  
1210 pseudorapidity constitute the jet-level inputs, with the track-level inputs listed in  
1211 Table 5.2. The track parameters and hit pattern are key indicators of whether or  
1212 not a track is fake. The FracRank variable is the ordered index of the tracks that  
1213 pass the ambiguity solver’s selection divided by the total number of successfully  
1214 reconstructed tracks in the event. The ambiguity solver processes track candidates  
1215 iteratively in order of an internal score (see Section 3.4.1), and the order in which  
1216 tracks are accepted is preserved. Since tracks with shared hits have lower scores,  
1217 tracks which do not require the removal of shared hits are likely to be processed  
1218 and accepted earlier on, whereas tracks with shared hits will be processed later and

---

<sup>2</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- $p_{\text{T}}$  radiation against which the  $t\bar{t}$  system recoils.

1219 potentially have their shared hits removed. Hence the FracRank variable gives an  
1220 indication of the of how easy it was for the track to be reconstructed.

| 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   |
| $r_{\text{first}}$ | Radius of first hit   |
| nDOF               | Number of degrees of freedom on the track                           |
| FracRank           | Ambiguity solver ordering variable                                  |

**Table 5.2:** Input features to the fake track classification NN. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [56], while split hits are hits used by multiple tracks which have been identified as merged, and therefore split.

1221 Track selection follows the loose selection described in Ref. [81] and outlined in  
1222 Table 5.3, which was found to improve the performance compared to previous tighter  
1223 selections, whilst ensuring good resolution of tracks and a low fake rate [56]. Inputs  
1224 are scaled to have a central value of zero and a variance of unity before training and  
1225 evaluation.

| 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 5.3:** 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 (see Section 3.2.2 for the IP definitions). Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [56]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

### 5.4.3 Model Hyperparameters

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

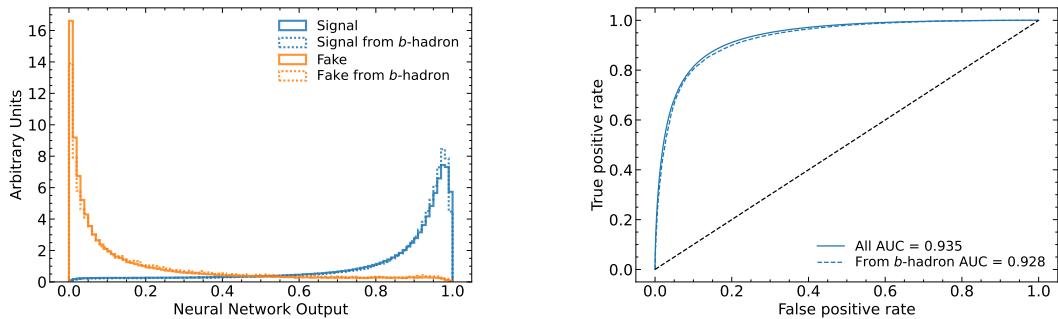
### 5.4.4 Results

In order to evaluate the fake track classification tool, a orthogonal test sample of 1 million tracks in jets in the combined  $t\bar{t}$  and  $Z'$  samples is used. The continuous scalar output from the NN model is interpreted as the probability that a given track is a signal track (i.e. not fake). Fig. 5.6 shows the performance of the fake track classification MVA. The signal and background classes are well separated in the output of the tool. Also shown is a receiver operating characteristic (ROC) curve,

| Hyperparameter        | Value  |
|-----------------------|--------|
| Batch size            | 2048   |
| Activation            | ReLU   |
| Optimiser             | Adam   |
| Initial learning rate | $1e-3$ |
| Training epochs       | 20     |
| Training tracks       | 40m    |
| Validation tracks     | 4m     |
| Testing tracks        | 4m     |

**Table 5.4:** Hyperparameter for the track classification model

which plots the rate of true positives against the rate of false positives over a scan of cut points on the NN output ranging from zero to one. The area under the curve (AUC) gives a summary of the aggregate classification power of the model. The fake track classification tool achieves an AUC of 0.935 for all tracks, which is indicative of a well-performing model. Considering only tracks from  $b$ -hadron decays, this value drops slightly to 0.928.



**Figure 5.6:** (left) Normalised histogram of the model output separated for signal and fake tracks, and further separated by those tracks which are from the decay of a  $b$ -hadron. (right) The ROC curve for all tracks (solid line) and tracks from the decay of a  $b$ -hadron (dashed line). The model is able to successfully discriminate between signal and fake tracks, and shows a very similar performance when looking specifically at tracks from the decay of a  $b$ -hadron.

Signal and fake track efficiencies at two different NN output cut points are shown in Table 5.5. The results demonstrate that the tool is effective in retaining 98.8% of signal tracks, while correctly identifying (and therefore enabling the removal of)

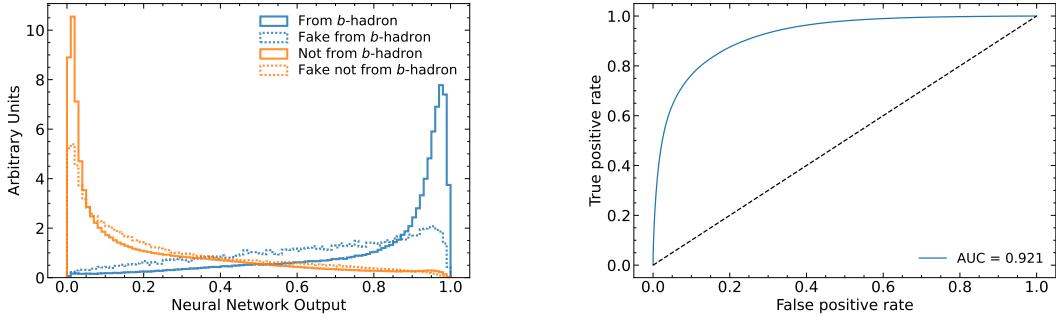
1252 45.6% of fake tracks. Table 5.5 also shows that a significant amount of tracks which  
1253 are labelled as both fake and from the decay of a  $b$ -hadron are also removed. This can  
1254 happen because fake tracks with  $\text{TMP} < 0.75$  are still matched to a truth particle,  
1255 which can be the decay product of a  $b$ -hadron.

| MVA<br>Output Cut | Signal Track Efficiency |          | Fake Track Efficiency |          |
|-------------------|-------------------------|----------|-----------------------|----------|
|                   | All                     | From $b$ | All                   | From $b$ |
| 0.06              | 98.8%                   | 98.9%    | 45.6%                 | 39.8%    |
| 0.12              | 97.3%                   | 97.5%    | 59.4%                 | 53.6%    |

**Table 5.5:** Good and fake track selection efficiencies for the combined  $t\bar{t}$  and  $Z'$  samples. Two working points are defined, cutting on the NN output at 0.06 and 0.12. The continuous output of the model allows for the tuning of good and fake track identification efficiencies.

## 1256 5.5 $b$ -hadron Track Identification

1257 After initial tests and investigation, it was found that fake tracks which were the  
1258 result of  $b$ -hadron decays actually aided  $b$ -tagging performance. The application of  
1259 a single tool which removed all fake tracks was therefore not optimal. A second  
1260 tool was therefore trained in the same manner as the first, this one was designed to  
1261 distinguish between those tracks which were from the decay of a  $b$ -hadron (FromB  
1262 and FromBC in Table 5.1) and those which were not (all other truth origins). The  
1263  $b$ -hadron decay track MVA was trained using the same setup as described above,  
1264 with the same tracks, input variables, and training procedure. The performance of  
1265 the model to separate  $b$ -hadron decay tracks from other tracks is shown in Fig. 5.7.  
1266 Using a selection WP of 0.1, the model can retain 98.5% of  $b$ -hadron tracks and  
1267 reject 46.2% of tracks not from the decay of a  $b$ -hadron. In Section 5.6, this model is  
1268 used in conjunction with the fake track identification MVA described in Section 5.4.4  
1269 to identify and remove fake tracks which are not from the decay of a  $b$ -hadron.



**Figure 5.7:** (left) Normalised histogram of the model output separated for tracks from the decay of a  $b$ -hadron and tracks from other sources. The two groups are further separated into those tracks which are fake. (right) The ROC curve for all tracks (solid line).

## 1270 5.6 Combined Approach

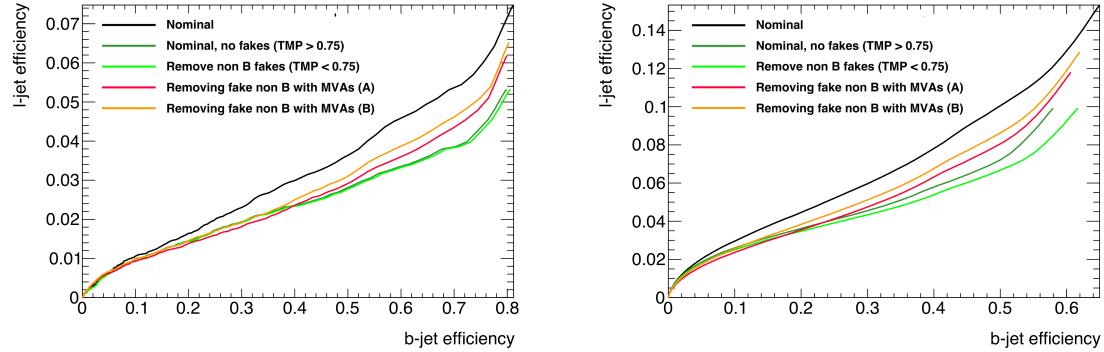
1271 A 2-dimensional cut was then used to only reject those tracks that had a high  
 1272 probability of being fake, and also a low probability of being a  $b$ -hadron decay track.

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

**Table 5.6:** Cut values for the fake and  $b$ -hadron decay track MVAs for the two defined working points. Working point “B” cuts more aggressively on the MVA outputs than WP “A”, removing more fake tracks but resulting in an increased loss of signal tracks (which here are all  $b$ -hadron decay tracks).

1273 The light-jet efficiency of SV1 is successfully reduced when using the combined tools  
 1274 to remove fake tracks that are not from a  $b$ -hadron decay, as shown in Fig. 5.8. At a  
 1275  $b$ -jet efficiency of 70%, the light-jet mistag rate for jets with  $250 < p_T < 400$  GeV  
 1276 is reduced from 0.054 to 0.044, a relative improvement of approximately 20%. For  
 1277 jets with  $400 < p_T < 1000$  GeV the mistage rate drops from 0.1 to 0.08 for a similar  
 1278 relative improvement of 20%. The performance of the fake track removal approach  
 1279 was also tested for the other low level vertexing algorithm: JetFitter. A similar level  
 1280 of improvement in the light-jet mistag rate was observed of up to a 20% reduction

for both low- and high- $p_T$  jets in the  $Z'$  sample. Together, these results demonstrate that by identifying and removing fake tracks which are not the result of the weak decay of a  $b$ -hadron, the performance of the low level tagging algorithms can be improved.



**Figure 5.8:** The effect of applying the fake track identification algorithm together with the  $b$ -hadron decay track identification on the jet tagging performance of SV1 for jets in the  $Z'$  sample with  $250 \text{ GeV} < p_T < 400 \text{ GeV}$  (left) and  $400 \text{ GeV} < p_T < 1 \text{ TeV}$  (right). The nominal SV1 light-jet rejection (black) is compared to two working points of fake track removal, labelled “A” (red) and “B” (orange), which represent two different 2D working points of the track classification tools. Removal of fake tracks based on truth information is shown by the green curves.

## 5.7 Conclusion

Fake tracks, which are prevalent in the core of high  $p_T$  jets, have an adverse impact on  $b$ -tagging performance. A ML tool to identify fake tracks has been developed, which can be used to limit the number of fake tracks being input to the  $b$ -tagging algorithms. Since it was found that  $b$ -hadron decay tracks can also be poorly reconstructed and thus marked as fake, it was deemed necessary also to train a second algorithm to detect  $b$ -hadron decay tracks so that the removal of these tracks could be avoided. Removing fake and non- $b$  decay tracks in this way was found to improve the light-jet mistagging rate of SV1 and JetFitter by up to 20% at high transverse momentum. The improvement achieved using the classification tools was in general comparable with the improvement achieved when using the truth information to remove the fake tracks not from the decay of a  $b$ -hadron.

**1297 Future Work**

1298 While removing tracks prior to their input to the low level tagging algorithms is  
1299 shown here to be beneficial, a more performant alternative might be to keep these  
1300 tracks but label them as being fake (for example using the output of the classification  
1301 tool), and allow the tagging algorithms to take this into consideration, potentially  
1302 making use of this information. This is not straightforward with manually optimised  
1303 taggers such as SV1 and JetFitter, but is possible with more advanced taggers as  
1304 described in Chapter 6.

1305 Tools which identify the origin of a given track have other potential uses. One  
1306 application is to isolate a relatively pure sample of fake tracks which can be used  
1307 to estimate the fake track rate in data, which would be useful for estimating the  
1308 uncertainty on fake track modelling. Another application would be to use the  
1309  $b$ -hadron track identification tool to improve the track-to-jet association. Both  
1310 applications are currently under investigation.

1311 The approach here works on a track-by-track basis, but a more sophisticated approach  
1312 would consider the correlations between the tracks inside a jet, as shown in Chapter 6.

1313 Also left for future work is to simultaneously train a single tool which discriminates  
1314 between all the truth origins listed in Table 5.1. Such a tool would be useful as a  
1315 general purpose multiclass classifier.

<sub>1316</sub> **Chapter 6**

<sub>1317</sub> **Graph Neural Network Flavour  
Tagger**

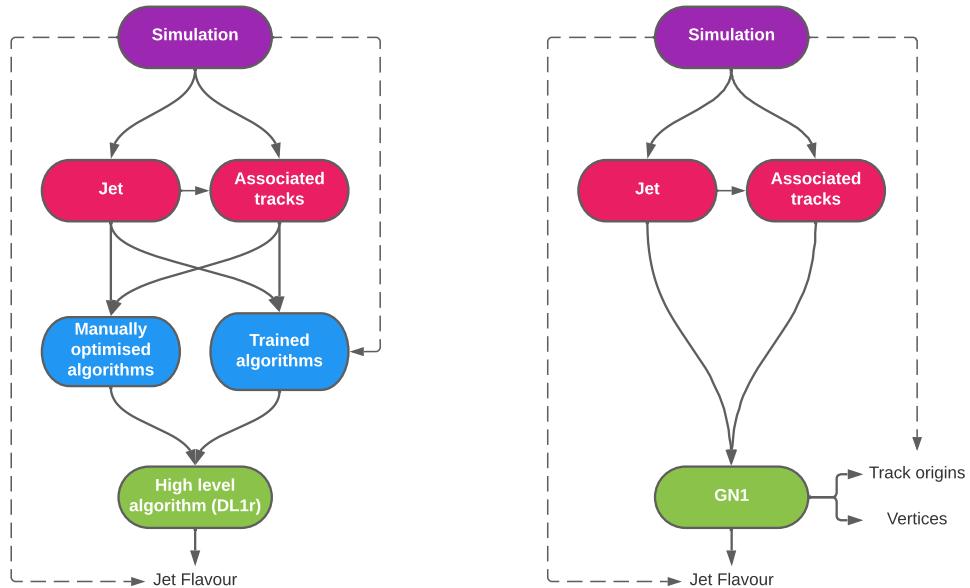
<sub>1319</sub> As discussed in Chapter 4, flavour tagging is the identification of jets instantiated  
<sub>1320</sub> from  $b$ - and  $c$ -hadrons. Flavour tagging is a critical component of the physics  
<sub>1321</sub> programme of the ATLAS experiment. It is of crucial importance for the study of the  
<sub>1322</sub> Standard Model (SM) Higgs boson and the top quark, which decay preferentially to  
<sub>1323</sub>  $b$ -quarks [112, 113], and additionally for several Beyond the Standard Model (BSM)  
<sub>1324</sub> resonances that readily decay to heavy flavour quarks [114].

<sub>1325</sub> This chapter introduces GN1, a novel ML-based flavour tagging algorithm based  
<sub>1326</sub> on graph neural networks (GNNs). In Section 6.1, an overview of the approach  
<sub>1327</sub> used for GN1 is provided. An introduction to the theory of GNNs is provided in  
<sub>1328</sub> Section 5.2. Details of the experimental setup are provided in Section 6.2, while  
<sub>1329</sub> the architecture of GN1 is specified in Section 6.3.3. In Section 6.3.4, the training  
<sub>1330</sub> procedure is described, and in Section 6.4 the results are shown.

<sub>1331</sub> **6.1 Motivation**

<sub>1332</sub> GN1 is a monolithic approach to flavour tagging as illustrated in Fig. 6.1. As opposed  
<sub>1333</sub> to the existing approach to flavour tagging described in Chapter 4, which relies  
<sub>1334</sub> on a two tiered approach requiring the use of both low- and high-level algorithms,  
<sub>1335</sub> GN1 takes as inputs information directly from an unordered variable number of  
<sub>1336</sub> tracks as input, and predicts the jet flavour without requiring outputs from the

intermediate low-level algorithms. In addition to predicting the flavour of the jet, the model predicts which physical processes produced the various tracks, and groups the tracks into vertices. These auxiliary training objectives provide valuable additional information about the contents of the jet and enhance the performance of the primary flavour prediction task. The use of GNNs offers a natural way to classify jets with variable numbers of unordered associated tracks (see Section 5.2), while allowing for the inclusion of auxiliary training objectives [115, 116].



**Figure 6.1:** Comparison of the existing flavour tagging scheme (left) and GN1 (right) [2]. The existing approach utilises low-level algorithms (shown in blue), the outputs of which are fed into a high-level algorithm (DL1r). Instead of being used to guide the design of the manually optimised algorithms, additional truth information from the simulation is now being used as auxiliary training targets for GN1. The solid lines represent reconstructed information, whereas the dashed lines represent truth information.

As described in Chapter 4, current flavour tagging algorithms utilise a two-tiered approach. The high-level tagger DL1r outputs variables which provides good discrimination between the different jet flavours. In contrast GN1 consists of only a single neural network, which takes tracks as inputs along with some kinematic information about the jet. As a result, it does not depend on the outputs of any other flavour tagging algorithm. A simple training of the model fully optimises its parameters, representing a significant simplification with respect to the optimisation procedure for

1351 DL1r. This is particularly important when optimising the tagger for new regions of  
1352 phase space (e.g.  $c$ -tagging or high- $p_T$   $b$ -tagging), or when the detector is upgraded  
1353 or the charged particle reconstruction or selection algorithms are re-optimised.

1354 GN1 is trained to learn about the internal structure of the jet through the use of two  
1355 auxiliary training objectives: the prediction of the underlying physics process from  
1356 which each track originated, and the grouping of tracks originating from a common  
1357 spatial position (i.e. a common vertex). These auxiliary objectives are meant to  
1358 guide the neural network towards a more complete understanding of the underlying  
1359 physics inside the jet, thereby removing the need for the low-level algorithms, which  
1360 previously contained information about the underlying physics in their design. The  
1361 training targets for the primary and auxiliary objectives are extracted from truth  
1362 information, i.e. information that is only available in simulation, as opposed to  
1363 reconstructed quantities available in both collision data and simulation.

1364 In this chapter, the following advantages of the GN1 approach will be demonstrated:

- 1365 1. GN1 boasts improved performance with respect to the current ATLAS flavour  
1366 tagging algorithms, with significantly larger background rejection rates for a  
1367 given signal efficiency. Alternatively the rejection rates can be kept fixed for a  
1368 substantial increase in signal efficiency, in particular at high- $p_T$ .
- 1369 2. The same network architecture can be easily optimised for a wider variety of  
1370 use cases (e.g.  $c$ -jet tagging and high- $p_T$  jet tagging) since there are no low-level  
1371 algorithms to retune.
- 1372 3. There are fewer algorithms to maintain.
- 1373 4. Alongside the network's prediction of the jet flavour, the auxiliary vertex and  
1374 track origin predictions provide more information on why a jet was (mis)tagged  
1375 or not. This information can also have uses in other applications, for instance  
1376 to explicitly reconstruct displaced decay vertices or to remove fake tracks.<sup>1</sup>

---

<sup>1</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. [56].

## 1377 6.2 Experimental Setup

### 1378 6.2.1 Datasets

1379 The datasets used to train the GN1 tagger are the same as described in Section 5.4.1.  
1380 The training dataset contains 30 million jets, 60% of which are  $t\bar{t}$  jets and 40%  
1381 of which are  $Z'$  jets. In order to evaluate the performance of the model during, a  
1382 statistically independent set of 500k testing jets from both the  $t\bar{t}$  and  $Z'$  samples are  
1383 used. For the final testing of the model and the creation of the performance plots,  
1384 a further 1 million independent testing jets from each of the  $t\bar{t}$  and  $Z'$  samples are  
1385 used. Before being fed into the model, the track- and jet-level inputs are normalised  
1386 to have a mean of zero and a variance of unity. The jet flavour labels are assigned  
1387 as described in Section 3.4.3. Truth labelled  $b$ -,  $c$ - and light-jets are kinematically  
1388 re-sampled in  $p_T$  and  $\eta$  to ensure identical distributions in these variables.

## 1389 6.3 Model Architecture

### 1390 6.3.1 Model Inputs

1391 As inputs, GN1 is fed two kinematic jet variables and an unordered set of up to 40  
1392 tracks which have been associated to the jet. Each track consists of 21 variables. The  
1393 kinematic jet variables are the jet transverse momentum and signed pseudorapidity.  
1394 The input variables which are provided for each track are listed in Table 6.1. For  
1395 each track, variables containing the track parameters and uncertainties, and detailed  
1396 information on the hit content are provided as inputs to the model.  
  
1397 In cores of high- $p_T$  jets, track density is high due to the increased multiplicity and  
1398 collimation of tracks (see Chapter 4). As a result, the separation between tracks  
1399 can be of the same order as the active sensor dimensions, resulting in an increase  
1400 in merged clusters and tracks which share hits [56]. Due to the relatively long  
1401 lifetimes of  $b$ -hadrons and  $c$ -hadrons, which can traverse several layers of the ID  
1402 before decaying and have highly collimated decay products, the presence of shared  
1403 or missing hits is a critical signature of heavy flavour jets.

1404 Dependence of the model on the absolute value of the azimuthal jet angle  $\phi$  is  
1405 explicitly removed by providing only the azimuthal angle of tracks relative to the jet  
1406 axis. The track pseudorapidity is also provided relative to the jet axis.

1407 Since heavy flavour hadrons can decay semileptonically approximately 20% of the time,  
1408 the presence of a reconstructed lepton in the jet carries discriminating information  
1409 about the jet flavour. To exploit this, a variant of GN1 called GN1Lep is trained in  
1410 addition to the baseline model. The GN1Lep variant is identical to the baseline model,  
1411 except for the inclusion an additional track-level input, leptonID, which indicates  
1412 if the track was used in the reconstruction of an electron, a muon or neither. The  
1413 variable is signed by the charge of the reconstructed lepton. The leptons used in the  
1414 definition of the leptonID variable are required to satisfy basic quality requirements.  
1415 The muons are required to be combined [117], and the electrons are required to pass  
1416 the *VeryLoose* likelihood-based identification working point [118].

1417 The selections applied to the tracks is the same as that used for the fake track  
1418 classification MVA described in Chapter 5. The full set of track selections is listed  
1419 in Table 5.3. This selection was found to improve the flavour tagging performance  
1420 compared to previous tighter selections, whilst ensuring good resolution of tracks  
1421 and a low fake rate [56]. However, Section 6.4.8 demonstrates that further relaxation  
1422 of the track selection requirements may be warranted.

1423 If more than 40 tracks are associated to a given jet, only the first 40 tracks with the  
1424 largest transverse IP significance<sup>2</sup>  $s(d_0)$  are fed into the model as inputs.

### 1425 6.3.2 Auxiliary Training Objectives

1426 In addition to the jet flavour classification, two auxiliary training objectives are  
1427 defined. The first auxiliary objective is the prediction of the physical process that  
1428 gave rise to each track within the jet (i.e. the track origin), while the second is the  
1429 prediction of track-pair vertex compatibility. Each auxiliary training objective comes  
1430 with a training target which, similar to the jet flavour label, is a truth labels derived

---

<sup>2</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 [38].

| Jet Input         | Description   |
|-------------------|---|
| $p_T$             | Jet transverse momentum   |
| $\eta$            | Signed jet pseudorapidity   |
| Track Input       | Description   |
| $q/p$             | Track charge divided by momentum                                    |
| $d\eta$           | 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 6.1:** Input features to the GN1 model [2]. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [56], while split hits are hits used on multiple tracks which have been identified as merged. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. The track leptonID is an additional input to the GN1Lep model.

1431 from the simulation. The presence of the auxiliary training objectives improves the  
1432 jet classification performance as demonstrated in Section 6.4.3.

1433 For the track origin prediction objective, each track is labelled with one of the  
1434 exclusive categories defined in Table 5.1 of Section 5.3 after analysing the particle  
1435 interaction (or lack thereof) which led to its formation. Since the presence of different  
1436 track origins is strongly related to the flavour of the jet, training GN1 to recognise  
1437 the origin of the tracks provides an additional handle on the classification of the  
1438 jet flavour. This task may also aid the jet flavour prediction by acting as a form of  
1439 supervised attention [119] - in detecting tracks from heavy flavour decays the model  
1440 may learn to pay more attention to these tracks.

1441 The vertexing auxiliary objective makes use of the fact that displaced decays of  $b$ -  
1442 and  $c$ -hadrons lead to secondary and tertiary vertices inside the jet, as described in  
1443 Section 4.1.1. The presence of displaced secondary vertices is not a completely clean  
1444 signal of a heavy flavour jet, as displaced secondary vertices can also occur in light-jets  
1445 as a result of material interactions, conversions, and long-lived particle decays (e.g.  
1446  $K_S^0$  and  $\Lambda^0$ ). For the auxiliary object, GN1 predicts a binary label for each pair of  
1447 tracks in the jet. The label has a value of 1 if the truth particles associated with the  
1448 two tracks in the pair originated from the same spatial point, and 0 otherwise. To  
1449 derive the corresponding truth labels for training, truth production vertices within 0.1  
1450 mm are merged. Track-pairs where one or both of the tracks in the pair have an origin  
1451 label of either Pileup or Fake are given a label of 0. Using the pairwise predictions  
1452 from the model, groups of tracks that have common compatibility can be formed,  
1453 resulting in the finding of vertices. Two existing low-level tagging algorithms, SV1  
1454 and JetFitter (introduced in Section 3.4.2), are currently used to find and reconstruct  
1455 vertices inside jets and are used as inputs to the existing jet flavour tagger DL1r.  
1456 The addition of this auxiliary training objective removes the need for inputs from a  
1457 dedicated secondary vertexing algorithm.

1458 Both of the auxiliary training objectives described here can be considered as “stepping  
1459 stones” on the way to classifying the flavour of the jet. By requiring the model to  
1460 predict the truth origin of each track and the vertex compatibility of each track-pair,  
1461 the model is guided to learn representations of the jet which are connected to the  
1462 underlying physics and therefore relevant for classifying the jet flavour.

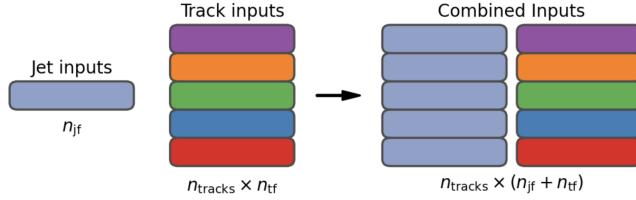
### 1463 6.3.3 Architecture

1464 As discussed in the previous sections, GN1 is a graph neural network which makes  
1465 use of auxiliary training objectives in order to determine the jet flavour. A coarse  
1466 optimisation of the network architecture hyperparameters (for example number of  
1467 layers and number of neurons per layer) has been carried out in order to maximise  
1468 the flavour tagging performance, but it is likely that further dedicated optimisation  
1469 studies could lead to further performance improvements.

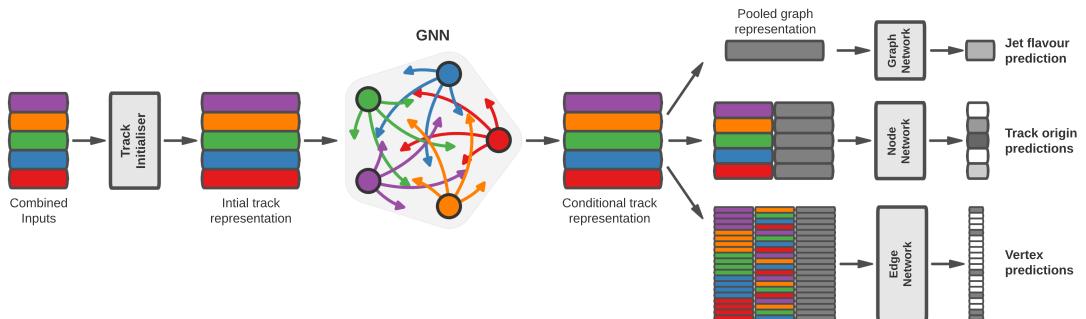
1470 The model architecture builds on a previous implementation of a GNN-based jet  
1471 tagger [116]. The previous approach was comprised of two separate graph neural  
1472 networks with the auxiliary tasks being performed at an intermediate stage after the  
1473 first and before the second. This two stage approach was found to be unnecessary and  
1474 as such GN1 simplifies the architecture into a single graph neural network with the  
1475 auxiliary tasks being performed at the end, alongside the primary jet classification  
1476 task. GN1 makes use of a more sophisticated graph neural network layer [120],  
1477 which is described in more detail below. The changes significantly improved tagging  
1478 performance and also led to a significant reduction in training time.

1479 As inputs, the model takes information about the jet and a number of associated  
1480 tracks, as detailed in Section 6.3.1. The jet variables are concatenated with the  
1481 variables for each track as shown in Fig. 6.2. The combined jet-track input vectors  
1482 are then fed into a per-track initialisation network with three hidden layers, each  
1483 containing 64 neurons, and an output layer with a size of 64, as shown in Fig. 6.3. The  
1484 track initialisation network is similar to a deep sets model [98], but does not include  
1485 a reduction operation (mean or summation) over the output track representations.  
1486 The initialisation network allows for initial per-track input processing without the  
1487 associated parameter count cost of the graph convolutional layers described below.

1488 The outputs of the track initialisation network are used to populate the nodes of a  
1489 fully connected graph, such that each node in the graph neighbours every other node.  
1490 Each node  $h_i$  in the graph corresponds to a single track in the jet, and is characterised  
1491 by a feature vector, also called a representation. The per-track output representations  
1492 from the initialisation networks are used as the initial feature vectors of each node  
1493 in the graph. In each layer of the graph network, output node representations  $h'_i$   
1494 are computed by aggregating the features of  $h_i$  and neighbouring nodes  $\mathcal{N}_i$  using  
1495 a multi-head attention mechanism ( $n = 2$ ) as described in Ref. [120, 121]. First,



**Figure 6.2:** 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 ( $n_{tf} = 21$ ) [2]. The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN1.



**Figure 6.3:** 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.

1496 the feature vectors of receiver and sender nodes are fed into two fully connected  
 1497 linear layers  $\mathbf{W}_r$  and  $\mathbf{W}_s$ , to produce an updated representation for each sender and  
 1498 receiver node  $\mathbf{W}_r h_i$  and  $\mathbf{W}_s h_j$ . These updated feature vectors are used to compute  
 1499 edge scores  $e(h_i, h_j)$  for each node pair, as in

$$e(h_i, h_j) = \mathbf{a} \cdot \theta [\mathbf{W}_r h_i + \mathbf{W}_s h_j], \quad (6.1)$$

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

$$a_{ij} = \text{softmax}_j [e(h_i, h_j)]. \quad (6.2)$$

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

$$h'_i = \sigma \left[ \sum_{j \in \mathcal{N}_i} a_{ij} \cdot \mathbf{W}_r h_j \right]. \quad (6.3)$$

1505 The set of operations described above constitute a single graph network layer. Three  
 1506 such layers are stacked to construct the graph network, representing a balance  
 1507 between achieving good performance in a reasonable time and avoiding overtraining  
 1508 due to inflation of the parameter count of the model. The final output from the graph  
 1509 neural network is a set of per-node (i.e. per-track) feature vectors that are conditional  
 1510 representations of each track given the other tracks in the jet. In order to perform  
 1511 the jet flavour prediction, a flattened global representation of the jet is needed. To  
 1512 produce this, the output track representations are combined using a weighted sum,  
 1513 where the weights are learned during training and therefore act as a form of attention  
 1514 over the different tracks. The flattened outputs from the sum are then fed into a  
 1515 fully connected feedforward neural network with four layers and three outputs, one

for each jet flavour. Two other separate fully connected feedforward neural networks are then also used to independently perform the auxiliary classification objectives of GN1. Both of the auxiliary classification tasks also take in the global representation of the jet as inputs. A summary of the different classification networks used for the various training objectives is shown in Table 6.2.

| Network                      | Hidden layers   | Output size | Label                    |
|------------------------------|-----------------|-------------|--------------------------|
| Node classification network  | 128, 64, 32     | 7           | Track origin             |
| Edge classification network  | 128, 64, 32     | 1           | Track-pair compatibility |
| Graph classification network | 128, 64, 32, 16 | 3           | Jet flavour              |

**Table 6.2:** A summary of GN1’s different classification networks used for the various training objectives, adapted from Ref. [2]. The hidden layers column contains a list specifying the number of neurons in each layer.

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

The edge classification network is used to predict whether the tracks in the track-pair belong to a common vertex. This network takes as inputs the concatenated representations from each pair of tracks and the global jet representation. Similar to the node network, the edge network has three hidden layers containing 128, 64 and 32 neurons respectively, and a single output, which is used to perform binary classification of the track-pair compatibility. The output predictions for the two auxiliary networks are used for the auxiliary training objectives discussed in Section 6.3.2.

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

### 1539 6.3.4 Training

1540 The full GN1 training procedure minimises the total loss function  $L_{\text{total}}$ , defined  
 1541 in Eq. (6.4). This loss is composed of three terms:  $L_{\text{jet}}$ , the categorical cross  
 1542 entropy loss over the different jet flavours;  $L_{\text{vertex}}$ , the binary track-pair compatibility  
 1543 cross entropy loss; and  $L_{\text{track}}$ , the categorical cross entropy loss for the track origin  
 1544 prediction.  $L_{\text{vertex}}$  is computed via a weighted average over all intra-jet track-pairs in  
 1545 the batch, and  $L_{\text{track}}$  is computed by a weighted average over all tracks in the batch,  
 1546 where the weights are described below.

$$L_{\text{total}} = L_{\text{jet}} + \alpha L_{\text{vertex}} + \beta L_{\text{track}} \quad (6.4)$$

1547 The different losses converge to different values during training, reflecting differences  
 1548 in the relative difficulty of the various training objectives. The values of  $L_{\text{vertex}}$  and  
 1549  $L_{\text{track}}$  are weighted by  $\alpha = 1.5$  and  $\beta = 0.5$  respectively to ensure they converge to  
 1550 similar values, giving them an equal weighting towards  $L_{\text{total}}$ . The values of  $\alpha$  and  $\beta$   
 1551 are chosen to ensure that  $L_{\text{jet}}$  converges to a larger value than either  $L_{\text{vertex}}$  and  $L_{\text{track}}$ ,  
 1552 which reflects the primary importance of the jet classification objective. It was found  
 1553 that in practice the overall performance of the model was not sensitive to modest  
 1554 changes in the loss weights  $\alpha$  and  $\beta$ . Pre-training using  $L_{\text{total}}$  (i.e. on all tasks) and  
 1555 fine tuning on only the jet classification task also did not improve performance versus  
 1556 the described standard setup, indicating that the auxiliary tasks are not in direct  
 1557 competition with the jet classification task. As there was a large variation in the  
 1558 relative abundance of tracks of the different origins, the contribution of each origin to  
 1559  $L_{\text{track}}$  was weighted by the inverse of the frequency of their occurrence. In vertexing  
 1560 loss  $L_{\text{vertex}}$ , the class weight for track-pairs where both tracks are from either a  $b$ - or  
 1561  $c$ -hadron was increased by a factor of two as compared with other track-pairs, to  
 1562 encourage the network to focus on correctly classifying heavy flavour vertices.

1563 GN1 can be trained with either the node or edge networks (and their corresponding  
 1564 auxiliary tasks), or both, removed, as discussed in Section 6.4.3. In such cases,  
 1565 the corresponding losses  $L_{\text{vertex}}$  and  $L_{\text{track}}$  are also removed from the calculation  
 1566 of the overall loss  $L_{\text{total}}$ . The performance of the resulting models provides a

1567 useful indication of the benefit of including the auxiliary tasks to the primary jet  
1568 classification objective.

1569 GN1 was trained for 100 epochs on 4 NVIDIA V100 GPUs. Each epoch takes  
1570 approximately 25 mins to complete over the training sample of 30 million jets  
1571 described in Section 6.2.1. The Adam optimiser [122] with an initial learning rate of  
1572  $1e-3$ , and a batch size of 4000 jets (spread across the 4 GPUs) was used. Typically  
1573 the validation loss, calculated on 500k jets, became stable after around 60 epochs.  
1574 The epoch that minimized the validation loss was used for evaluation. GN1 has  
1575 been integrated into the ATLAS software [53] using ONNX [123]. The test sample  
1576 jet flavour predictions scores are computed using the ATLAS software stack as a  
1577 verification of this process.

## 1578 6.4 Results

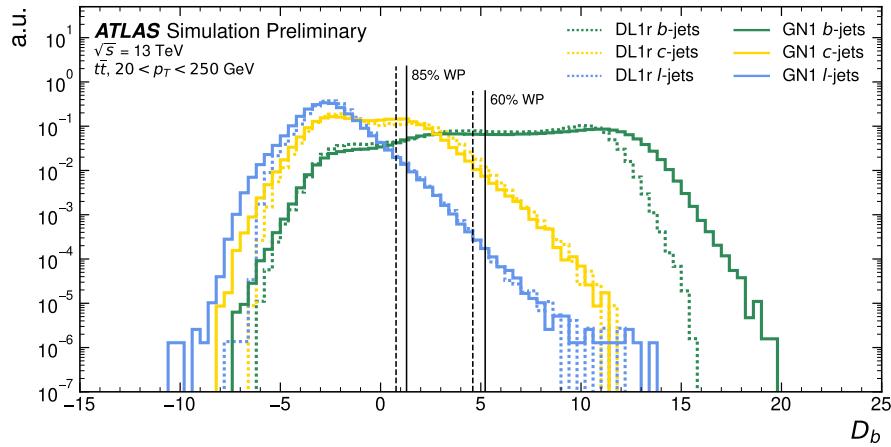
1579 The GN1 tagger is evaluated both as a  $b$ -tagging and  $c$ -tagging algorithm in Sec-  
1580 tion 6.4.1 and Section 6.4.2 respectively. Evaluation is performed separately on  
1581 both jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV and jets in the  $Z'$  sample with  
1582  $250 < p_T < 5000$  GeV. The performance of the model is compared to the DL1r  
1583 tagger [78, 124], which has been retrained on 75 million jets from the same samples  
1584 as GN1. The input RNNIP tagger [80] to DL1r has not been retrained. As discussed,  
1585 each tagger predicts the probability that a jet belongs to the  $b$ -,  $c$ - and light-classes.  
1586 To use the model for  $b$ -tagging, these probabilities are combined into a single score  
1587  $D_b$ , which is defined as

$$D_b = \log \frac{p_b}{(1 - f_c)p_l + f_c p_c}, \quad (6.5)$$

1588 where  $f_c$  is a free parameter that determines the relative weight of  $p_c$  to  $p_l$  in the  
1589 score  $D_b$ , controlling the trade-off between  $c$ - and light-jet rejection performance.  
1590 The choice of  $f_c$  is arbitrary, and is optimised based upon the desired light- vs  $c$ -jet  
1591 rejection performance. This parameter is set to a value of  $f_c = 0.018$  for the DL1r  
1592 model, obtained through an optimisation procedure described in Ref. [78]. Based on  
1593 a similar optimisation procedure, a value of  $f_c = 0.05$  is used for the GN1 models.

1594 A fixed-cut working point (WP) defines the corresponding selection applied to the  
 1595 tagging discriminant  $D_b$  in order to achieve a given efficiency on the inclusive  $t\bar{t}$   
 1596 sample.

1597 A comparison of the  $b$ -tagging discriminant  $D_b$  between DL1r and GN1 is shown in  
 1598 Fig. 6.4. The shapes of the  $D_b$  distributions are generally similar for  $b$ -,  $c$ - and light-  
 1599 jets between both models, however, GN1 shifts the  $b$ -jet distribution to higher values  
 1600 of  $D_b$  in the regions with the greatest discrimination. The GN1  $c$ -jet distribution is  
 1601 also shifted to lower values of  $D_b$  when compared with DL1r, enhancing the separation  
 1602 and indicating that GN1 is improving  $c$ -jet rejection when compared with DL1r.



**Figure 6.4:** Comparison between the DL1r and GN1  $b$ -tagging discriminant  $D_b$  for jets in the  $t\bar{t}$  sample [2]. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1. The distributions of the different jet flavours have been normalised to unity area.

#### 1603 6.4.1 $b$ -tagging Performance

1604 The performance of  $b$ -tagging algorithms is quantified by their ability to reject  $c$ - and  
 1605 light-jets for a given  $b$ -jet selection efficiency WP. In order to compare the  $b$ -tagging  
 1606 performance of the different taggers for the  $b$ -jet tagging efficiencies in the range  
 1607 typically used by analyses, the corresponding  $c$ - and light-jet rejection rates are  
 1608 displayed in Figs. 6.5 and 6.6 for jets in the  $t\bar{t}$  and  $Z'$  samples respectively. Four  
 1609 standard WPs are defined with  $b$ -jet tagging efficiencies of 60%, 70%, 77% and 85%

1610 respectively. These WPs are commonly used by physics analyses depending on their  
1611 specific signal and background requirements. The WPs are defined based on jets  
1612 in the  $t\bar{t}$  sample only. Due to the much higher jet  $p_T$  range in the  $Z'$  sample, and  
1613 the increased difficulty in tagging jets at high- $p_T$  (see Chapter 4), the  $b$ -jet tagging  
1614 efficiencies for jets in the  $Z'$  sample are lower than the corresponding WPs calculated  
1615 in the  $t\bar{t}$  sample. For instance the WP cut value computed to provide a 70%  $b$ -jet  
1616 tagging efficiency on the  $t\bar{t}$  sample results in a  $b$ -jet tagging efficiency of just  $\sim$ 30%  
1617 on the  $Z'$  sample. In order to account for this, the range of  $b$ -jet tagging efficiencies  
1618 displayed for plots showing the performance for jets in the  $Z'$  sample (for example  
1619 Fig. 6.6) is chosen to span the lower efficiencies achieved in the  $Z'$  sample at high- $p_T$ .

1620 For jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV, GN1 demonstrates considerably  
1621 better  $c$ - and light-jet rejection when compared with DL1r across the full range of  
1622  $b$ -jet tagging efficiencies studied. The relative improvement is strongly dependent  
1623 on the  $b$ -jet tagging efficiency under study. The largest improvements are found at  
1624 lower  $b$ -jet tagging efficiencies. At a  $b$ -jet tagging efficiency of 70%, the  $c$ -jet rejection  
1625 improves by a factor of  $\sim$ 2.1 while the light-jet rejection improves by a factor of  $\sim$ 1.8  
1626 with respect to DL1r. For high- $p_T$  jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV,  
1627 GN1 also brings a significant performance improvement with respect to DL1r across  
1628 the range of  $b$ -jet tagging efficiencies studied. Again, the largest relative improvement  
1629 in performance comes at the lower  $b$ -jet tagging efficiencies. At a  $b$ -jet efficiency of  
1630 30%, GN1 improves the  $c$ -jet rejection with respect to DL1r by a factor of  $\sim$ 2.8 and  
1631 the light-jet rejection by a factor of  $\sim$ 6. The performance comparison at lower  $b$ -jet  
1632 tagging efficiencies is made more difficult due to the increased statistical uncertainties  
1633 which result from the high rejection of background. Given that GN1 exploits the  
1634 low-level detector information in a more complete and sophisticated way than DL1r,  
1635 further studies are needed to confirm that the performance gain observed in these  
1636 simulated samples is also observed in experimental data.

1637 The GN1Lep variant of GN1 demonstrates further improved performance with respect  
1638 to the baseline model. This demonstrates the additional jet flavour discrimination  
1639 power provided by the leptonID track input. For jets in the  $t\bar{t}$  sample, the relative  $c$ -  
1640 jet rejection improvement with respect to GN1 at the 70%  $b$ -jet WP is approximately  
1641 25%. The improvement in light-jet rejection also increases by 40% at the same WP.  
1642 For jets in the  $Z'$  sample, the relative  $c$ -jet rejection (light-jet rejection) performance

1643 with respect to GN1 improves by approximately 10% (25%) at a  $b$ -jet tagging  
1644 efficiency of 30%.

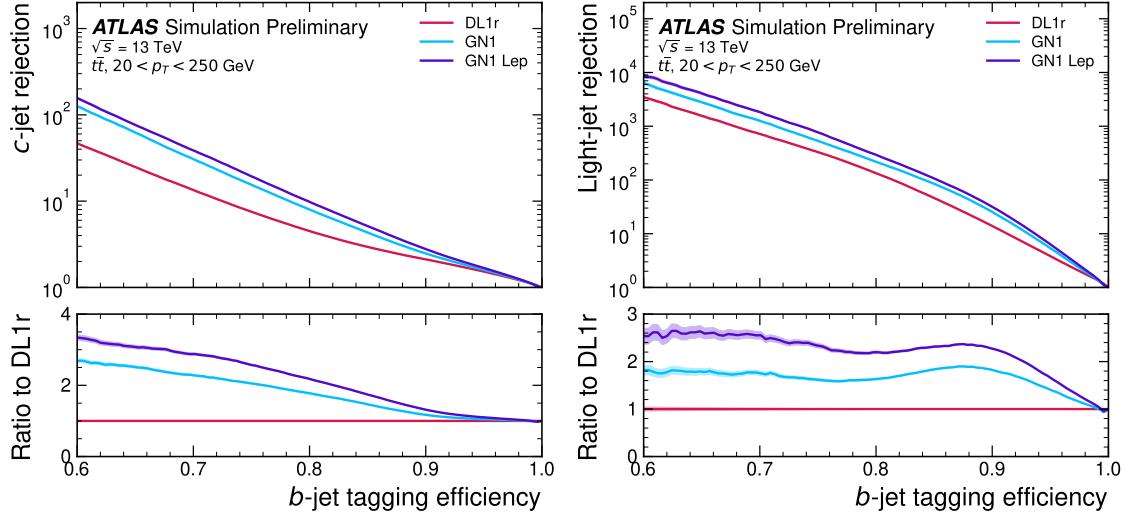
1645 In general, the performance of all the taggers is strongly dependent on the jet  $p_T$ .  
1646 This is due to the increased multiplicity and collimation of tracks, and the displaced  
1647 decays that result from within the heavy flavour jets (see Chapter 4). Together,  
1648 they contribute to a reduced reconstruction efficiency for heavy flavour tracks, and a  
1649 general degradation in quality of tracks inside the core of a jet, which in turn reduces  
1650 the jet tagging performance.

1651 In order to study how the tagging performance changes as a function of the jet  $p_T$ ,  
1652 the  $b$ -jet tagging efficiency as a function of  $p_T$  for a fixed light-jet rejection of 100 in  
1653 each bin is shown in Fig. 6.7. For jets in the  $t\bar{t}$  sample, at a fixed light-jet rejection  
1654 of 100, GN1 improves the  $b$ -jet tagging efficiency by approximately 4% across all the  
1655 jet  $p_T$  bins. Meanwhile, GN1Lep again demonstrates improved performance with  
1656 respect to GN1, in particular at lower  $p_T$ . The relative increase in the  $b$ -jet tagging  
1657 efficiency increases from 4% to 8% with respect to DL1r. For jets in the  $Z'$  sample,  
1658 GN1 again outperforms DL1r across the entire jet  $p_T$  range studied. The largest  
1659 relative improvement in performance is found at the highest transverse momenta  
1660 of jet  $p_T > 2 \text{ TeV}$ , and corresponds to an approximate factor of 2 improvement in  
1661 efficiency with respect to DL1r.

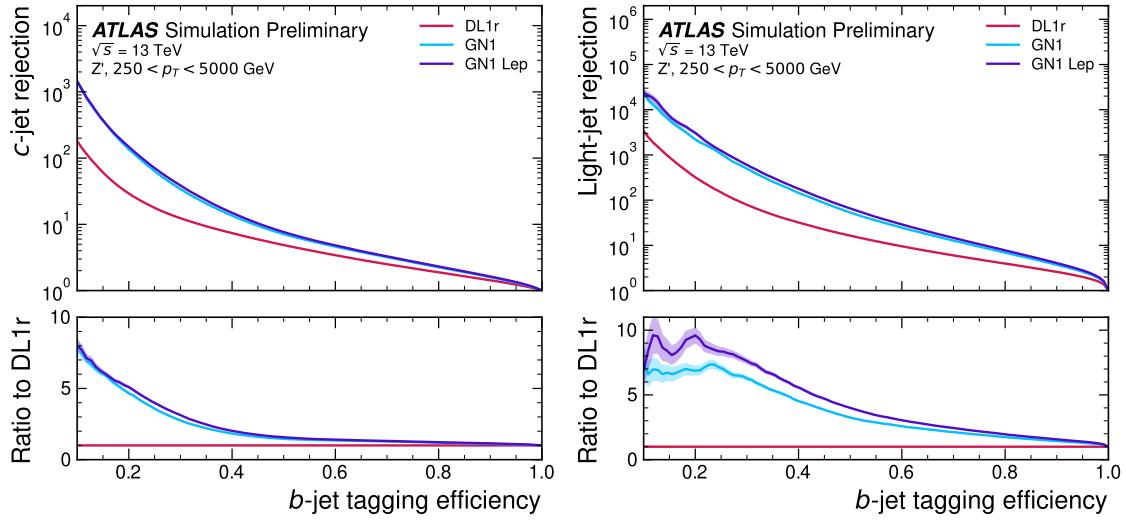
1662 The performance of the model was also evaluated as a function of the average  
1663 number of pileup interactions in the event. No significant dependence of the tagging  
1664 performance was observed.

#### 1665 6.4.2 $c$ -tagging Performance

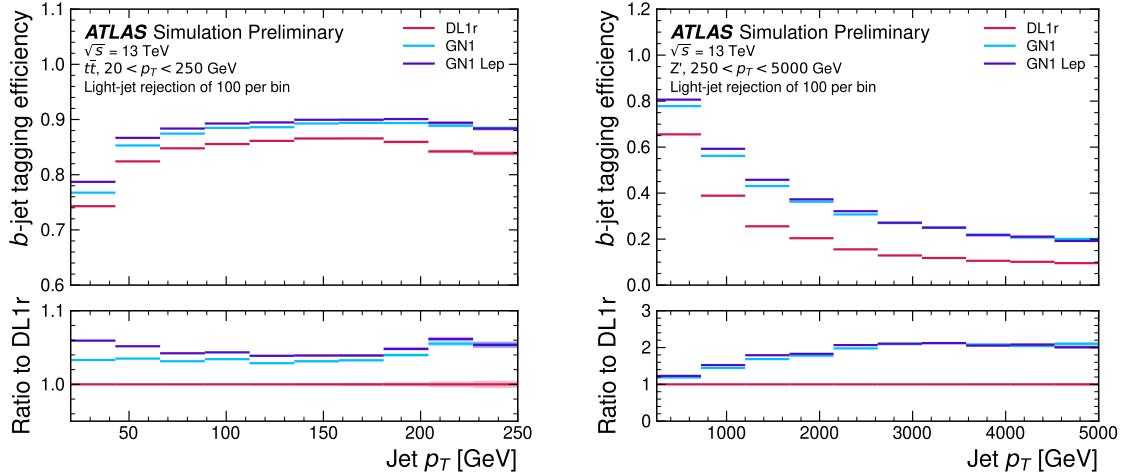
1666 As discussed previously, GN1 does not rely on any inputs from manually optimised  
1667 low-level tagging algorithms. Since these algorithms were originally designed and  
1668 tuned with the aim of  $b$ -tagging, and not  $c$ -tagging, the low level tagging algorithms  
1669 may perform suboptimally for  $c$ -tagging purposes. The tagging of  $c$ -jets therefore  
1670 presents a compelling use case for GN1. As each of the the models is trained with  
1671 three output classes, using it as a  $c$ -tagging algorithm is trivially analogous to the  
1672 approach used for  $b$ -tagging. The model output probabilities are combined into a  
1673 single score  $D_c$ , which is defined similarly to Eq. (6.5) as



**Figure 6.5:** 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  $20 < p_T < 250 \text{ GeV}$  [2]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $b$ -jet tagging efficiencies usually probed in these regions of phase space.



**Figure 6.6:** 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  $250 < p_T < 5000 \text{ GeV}$  [2]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $b$ -jet tagging efficiencies usually probed in these regions of phase space.



**Figure 6.7:** The  $b$ -jet tagging efficiency for jets in the  $t\bar{t}$  sample (left) and jets in the  $Z'$  sample (right) as a function of jet  $p_T$  with a fixed light-jet rejection of 100 in each bin [2]. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1 and GN1Lep. GN1 demonstrates improved performance with respect to DL1r across the  $p_T$  range shown. Binomial error bands are denoted by the shaded regions.

$$D_c = \log \frac{p_c}{(1 - f_b)p_l + f_b p_b}. \quad (6.6)$$

A value of  $f_b = 0.2$  is used for all models, based on the same optimisation procedure that was used for the  $b$ -tagging use case. Similar to Section 6.4.1, the different taggers are compared to one another by scanning through a range of  $c$ -jet tagging efficiencies and plotting the corresponding  $b$ - and light-jet rejection rates. As in Section 6.4.1, the WPs are defined using jets in the  $t\bar{t}$  sample. Standard  $c$ -jet tagging efficiency WPs used by physics analyses are significantly lower than the  $b$ -tagging WPs in order to maintain reasonable  $b$ - and light-jet rejection rates. This is reflected in the range of  $c$ -jet tagging efficiencies used in  $c$ -tagging plots such as Figs. 6.8 and 6.9. Fig. 6.8 displays the  $c$ -tagging performance of the models on the jets in the  $t\bar{t}$  sample. GN1 is shown to perform significantly better than DL1r. Similar to the  $b$ -tagging case, the  $b$ - and light-jet rejection improve most at lower  $c$ -jet tagging efficiencies, with the  $c$ -jet rejection (light-jet rejection) improving by a factor 2 (1.6) with respect to DL1r at a  $c$ -jet tagging efficiency of 25%. GN1Lep again outperforms GN1, though the improvements are more modest than observed for the  $b$ -tagging use

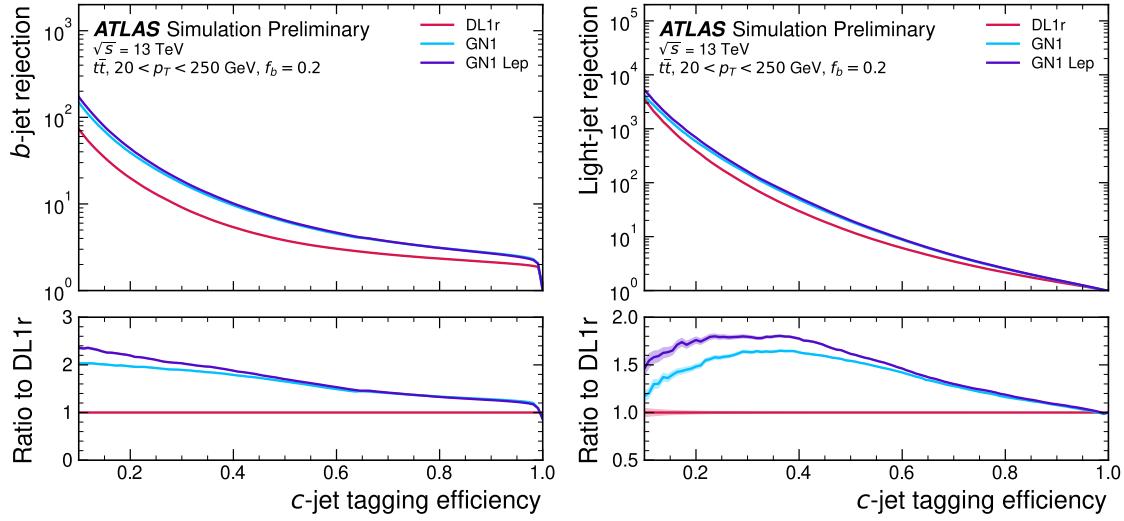
1688 case, with the both the  $b$ -jet rejection (light-jet rejection) improving with respect to  
1689 GN1 by approximately 10% (20%) at the 25%  $c$ -jet WP. Fig. 6.9 shows the  $c$ -tagging  
1690 performance on the jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV. Both GN1 and  
1691 GN1Lep perform similarly, improving the  $b$ -jet rejection by 60% and the light-jet  
1692 rejection by a factor of 2 at the 25%  $c$ -jet WP.

### 1693 6.4.3 Ablations

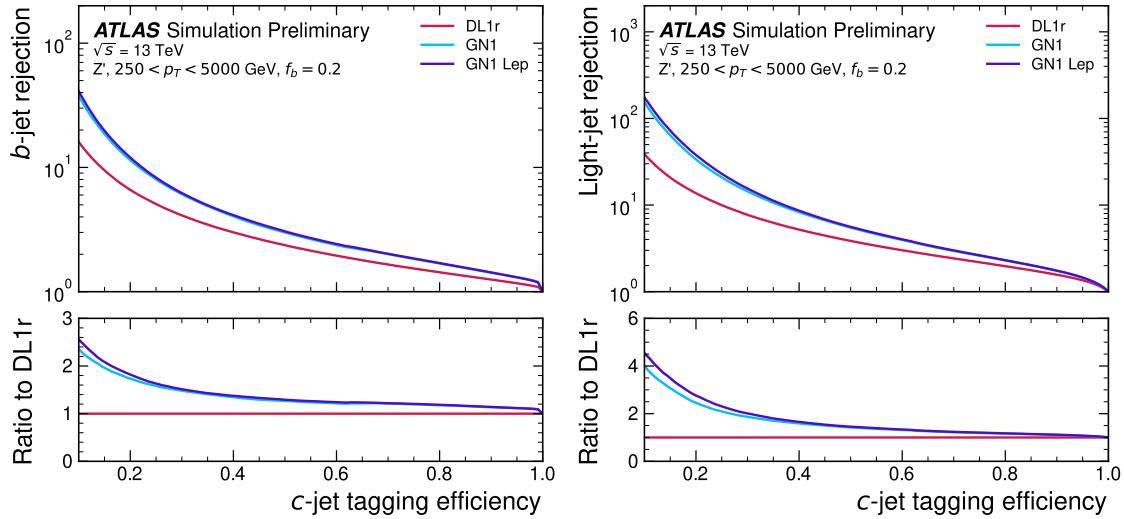
1694 Ablation studies (the removal of certain components of a given model in order to  
1695 study the impact of that component) are carried are carried out to determine the  
1696 importance of the auxiliary training objectives of GN1 to the overall performance.  
1697 The “GN1 No Aux” variant retains the primary jet classification objective, but  
1698 removes both track classification and vertexing auxiliary objectives (see Section 6.3.2)  
1699 and correspondingly only minimises the jet classification loss. The “GN1 TC” variant  
1700 includes track classification objective but not the vertexing objective. Finally, the  
1701 “GN1 Vert” includes the vertexing objective, but not the track classification objective.

1702 For jets in both the  $t\bar{t}$  and  $Z'$  samples, a general trend is observed that the models  
1703 trained without one or both of the auxiliary objectives results in significantly reduced  
1704  $c$ - and light-jet rejection when compared with the baseline GN1 model. This result  
1705 is shown clearly in Figs. 6.10 and 6.11. For jets in the  $t\bar{t}$  sample, the performance of  
1706 GN1 No Aux is similar to DL1r, while GN1 TC and GN1 Vert perform similarly to  
1707 each other. For jets in the  $Z'$  sample meanwhile, the GN1 No Aux model already  
1708 shows a clear improvement in  $c$ - and light-jet rejection when compared with DL1r at  
1709 lower  $b$ -jet tagging efficiencies. Similar to jets in the  $t\bar{t}$  sample, GN1 TC and GN1  
1710 Vert perform similarly, and bring large gains in background rejection when compared  
1711 with GN1 No Aux, but the combination of both auxiliary objectives yields the best  
1712 performance.

1713 It is notable that the GN1 No Aux model matches or exceeds the performance of  
1714 DL1r without the need for inputs from the low-level algorithms. This indicates that  
1715 the performance improvements enabled by the improved neural network architecture  
1716 used in GN1 appear to be able to compensate for the removal of the low-level  
1717 algorithm inputs. The GN1 TC and GN1 Vert variants each similarly outperform  
1718 DL1r, demonstrating that both contribute to the overall high performance of the  
1719 baseline model. The overall best performing model is the full version of GN1 trained



**Figure 6.8:** 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 [2]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $c$ -jet tagging efficiencies usually probed in these regions of phase space.



**Figure 6.9:** 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 [2]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The  $x$ -axis range is chosen to display the  $c$ -jet tagging efficiencies usually probed in these regions of phase space.

1720 with both auxiliary objective, demonstrating that the two auxiliary objectives are  
1721 complementary.

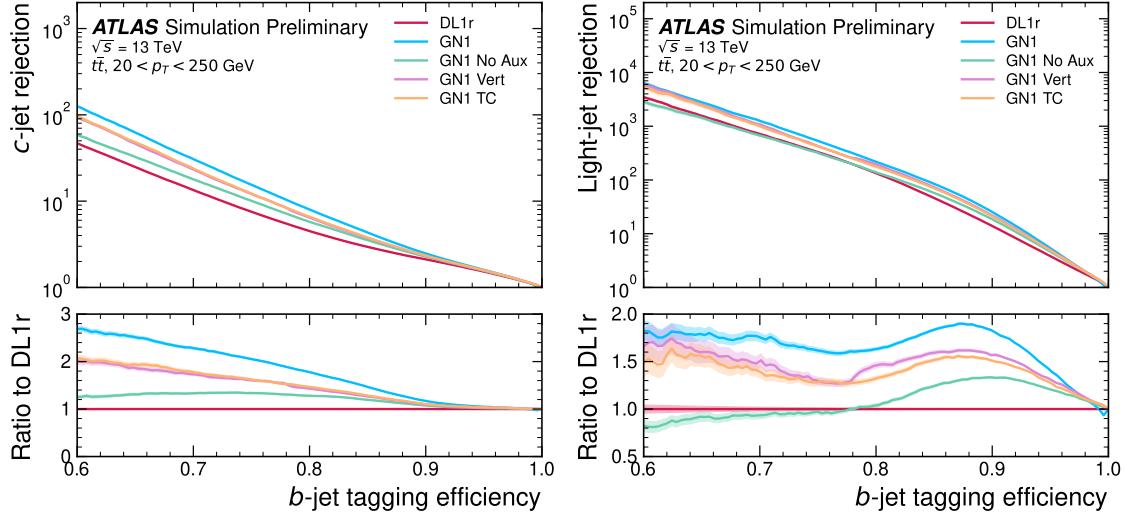
#### 1722 6.4.4 Inclusion of Low-Level Vertexing Algorithms

1723 As already mentioned, GN1 does not include any inputs from the low-level tagging  
1724 algorithms, including the vertexing algorithms SV1 and JetFitter [61]. Since these  
1725 algorithms are known to play a key role in contributing to the performance of DL1r, it  
1726 was necessary to study whether their inclusion in GN1 resulted in further performance  
1727 improvements. In a dedicated training of GN1 the SV1 and JetFitter tagger outputs  
1728 were added to the GN1 jet classification network as an input, similar to how they  
1729 are used in DL1r. These outputs include information on the reconstructed vertices,  
1730 including the number of vertices, and the properties of the highest ranking recon-  
1731 structed vertex (in the case of JetFitter). In addition, the index of the reconstructed  
1732 SV1 or JetFitter vertices were included as two track-level inputs to GN1. These  
1733 indices were also used to construct an input feature for the edge classification  
1734 network used to identify vertices, which was given a value of one if the track-pair  
1735 were from a common reconstructed SV1 or JetFitter vertex, and zero otherwise. The  
1736 jet classification performance of this GN1 model was not significantly different to the  
1737 baseline model, and in some cases the performance was slightly reduced. GN1 does  
1738 not benefit from the inclusion of information from SV1 and JetFitter, indicating  
1739 that the model is able to reconstruct the relevant information provided by these  
1740 low-level algorithms. The study also demonstrates that the model can function as a  
1741 highly performant standalone tagger that does not require (beyond retraining) any  
1742 manual optimisation to achieve good performance in a wide range of phase spaces.  
1743 A dedicated look at the vertexing performance of GN1 with some comparisons to  
1744 SV1 and JetFitter is found in Section 6.4.6

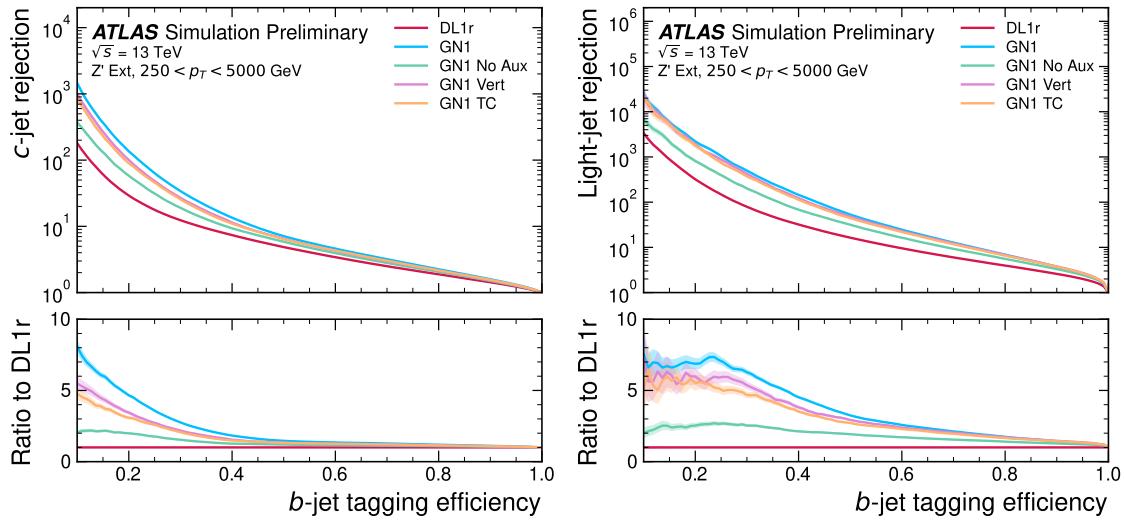
determine  
ranking

#### 1745 6.4.5 Jet Display Diagrams

1746 The auxiliary training objectives of GN1 allow for improved model interpretability,  
1747 which is especially important for a monolithic approach as the low level taggers,  
1748 which provide useful physical insight, are no longer present. Figs. 6.12 and 6.13  
1749 provide example comparisons of the true origin and vertexing information compared



**Figure 6.10:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $t\bar{t}$  jets with  $20 < p_T < 250$  GeV, for the nominal GN1, in addition to configurations where no (GN1 No Aux), only the track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary objectives are deployed [2]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1.



**Figure 6.11:** 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 [2]. The ratio to the performance of the DL1r algorithm is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1r and  $f_c = 0.05$  is used for GN1.

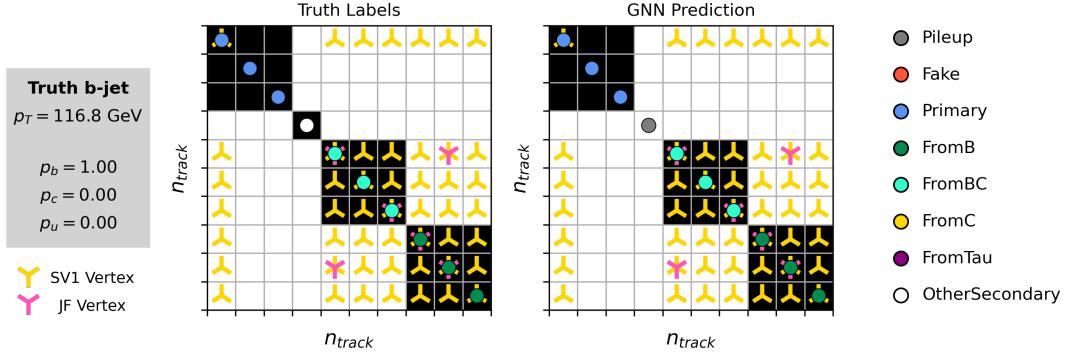
1750 with the predicted values from GN1, SV1 and JetFitter. Such comparisons can be  
 1751 used to provide an indication that GN1 reconstructs the correct representation of the  
 1752 jet structure, and may also help to identify limitations of the model. In the figures,  
 1753 the tracks in the jet are indexed twice on each of the  $x$ - and  $y$ -axes, and tracks are  
 1754 grouped into vertices along with other tracks as indicated by common markings in  
 1755 the relevant rows and columns.

1756 In Fig. 6.12, GN1 correctly groups the three primary tracks as having come from the  
 1757 primary vertex. The  $b$ -hadron and  $b \rightarrow c$ -hadron decay vertices are also correctly  
 1758 predicted, and the origin of the tracks in each is correct. There is a single OtherSec-  
 1759 ondary track which GN1 incorrectly predicts as having come pileup. Meanwhile SV1  
 1760 (by design) merges the two heavy flavour decay vertices, but incorrectly includes a  
 1761 track from the primary vertex. JetFitter reconstructs two vertices, one which is a  
 1762 combination of two tracks from different truth vertices and two other single track  
 1763 vertices in each of the heavy flavour vertices. GN1 also predicts the flavour of the jet  
 1764 with a high degree of certainty.

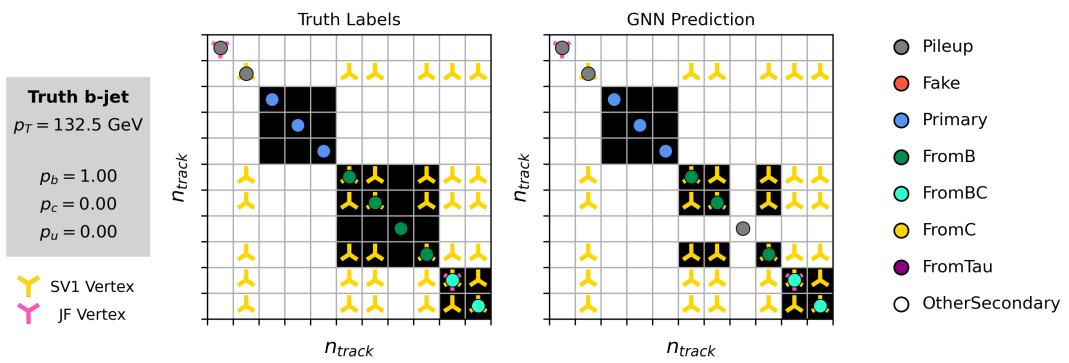
1765 Similarly, Fig. 6.12 shows that GN1 is able to relatively accurately predict the origin  
 1766 and vertex information of tracks inside a jet. The pileup tracks and primary vertex  
 1767 tracks are correctly identified, and the heavy flavour decay tracks are also correctly  
 1768 identified with the exception of one of the  $b$ -hadron decay tracks. Again, SV1 merges  
 1769 the two heavy flavour decay vertices along with a track from pileup, while JetFitter  
 1770 shows signs of being underconstrained by reconstructing two single track vertices,  
 1771 one with a pileup track and one with a track from a  $b \rightarrow c$ -hadron decay.

#### 1772 6.4.6 Vertexing Performance

1773 From the track-pair vertex prediction described in Section 6.3.2, tracks can be  
 1774 partitioned into compatible groups representing vertices (see [116]). As such, GN1  
 1775 can perform vertex “finding”, but not vertex “fitting”, i.e. the reconstruction of  
 1776 a vertex’s properties, which currently still requires the use of a dedicated vertex  
 1777 fitter. In order to study the performance of the different vertexing tools, the truth  
 1778 vertex label of the tracks, discussed in Section 6.3.2, are used. To estimate the  
 1779 efficiency with which GN1 manages to find vertices inclusively, vertices containing  
 1780 tracks identified as coming from a  $b$ -hadron are merged together and compared to the  
 1781 inclusive truth decay vertices that result from a  $b$ -hadron decay (where if there are



**Figure 6.12:** A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a  $b$ -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities  $p_b$ ,  $p_c$  and  $p_u$  are rounded to three significant figures. GN1 improves over SV1 and JetFitter by successfully determining the origins and vertex compatibility of all the tracks in the jet with the exception of the truth OtherSecondary track, which is misidentified as pileup.



**Figure 6.13:** A schematic showing the truth track origin and vertex information (left) compared with the predictions from GN1 (right). The diagrams show the truth (left) and predicted (right) structure of a  $b$ -jet. The shaded black boxes show the groupings of tracks into different vertices. Vertices reconstructed by SV1 and JetFitter are also marked. Class probabilities  $p_b$ ,  $p_c$  and  $p_u$  are rounded to three significant figures. GN1 improves over SV1 and JetFitter by successfully determining the origins and vertex compatibility of all but one tracks in the jet.

multiple distinct truth vertices from a  $b$ -hadron decay they are also merged together). Vertices are compared with the target truth vertex and the number of correctly and incorrectly assigned tracks is computed. Since secondary vertex information is only recovered for reconstructed tracks, a vertex finding efficiency of 100% denotes that all possible secondary vertices are found given the limits set by the track reconstruction efficiency. A vertex is considered matched if it contains at least 65% of the tracks in the corresponding truth vertex, and has a purity of at least 50%. GN1 manages to achieve an inclusive reconstruction efficiency in  $b$ -jets of  $\sim 80\%$ , demonstrating that it effectively manages to identify the displaced vertices from  $b$ -hadron decays.

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

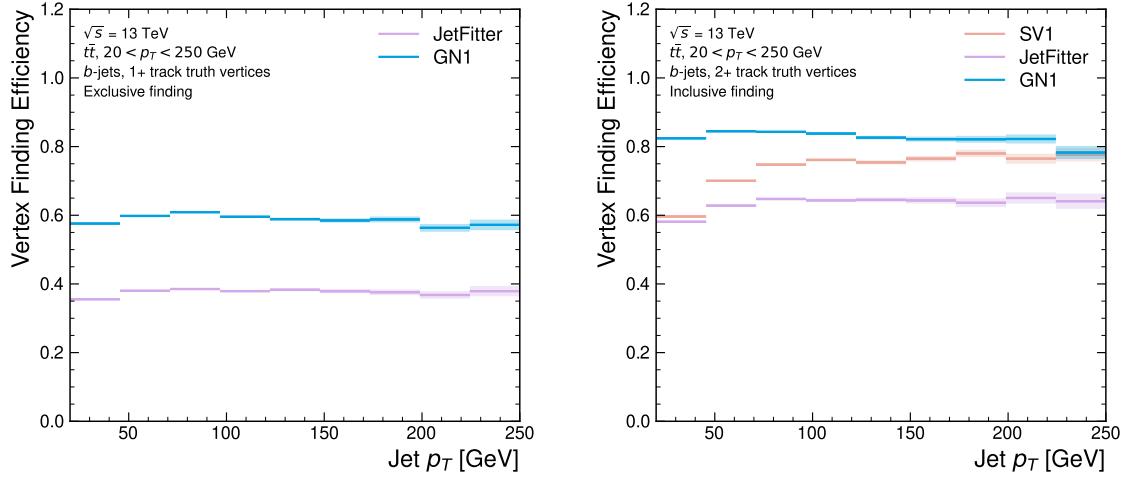
The starting point for the secondary vertex finding efficiency in both the exclusive and inclusive cases is to select truth secondary vertices, defined as those containing only inclusive  $b$ -hadron decays. For exclusive vertex finding, these truth secondary vertices can be used directly as the denominator for the efficiency calculation. Meanwhile for the inclusive efficiency all such truth secondary vertices in the jet are merged into a single inclusive target vertex. Correspondingly, for the inclusive vertex finding case, the vertices found by JetFitter are merged into a single vertex, and the vertices found by GN1, which contain at least one predicted  $b$ -hadron decay track, are also

merged similarly. SV1 does not require any vertex merging. Only jets containing a single  $b$ -hadron at truth level are considered.

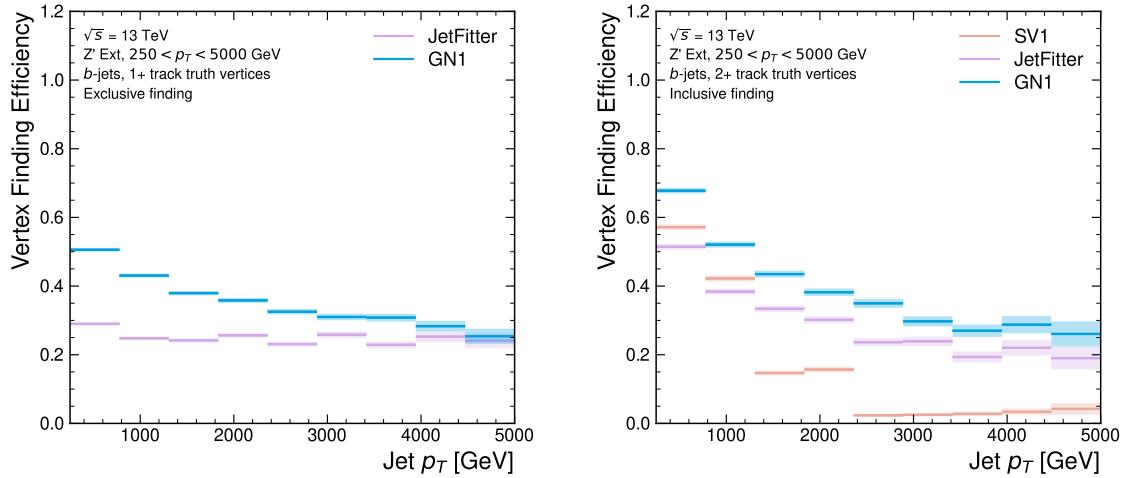
Next, vertices in the jet found by the different vertexing tools are compared with the target truth vertices. The number of correctly and incorrectly assigned tracks is computed. In order to call a vertex efficient, it is required to contain at least 65% of the tracks in the corresponding truth vertex,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

Vertex finding efficiencies for  $b$ -jets in the  $t\bar{t}$  sample are displayed as a function of  $p_T$  separately for the inclusive and exclusive approaches in Fig. 6.14. For  $b$ -jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV, the exclusive vertex finding efficiency of JetFitter and GN1 is relatively flat as a function of  $p_T$ . For the truth secondary vertices in this  $p_T$  region, JetFitter efficiently finds approximately 40% and GN1 finds approximately 55%. When finding vertices inclusively the vertex finding efficiency is generally higher. An increased dependence on  $p_T$  is also visible for JetFitter and SV1. As the jet  $p_T$  increases from 20 GeV to 100 GeV, the efficiency of JetFitter increases from 60% to 65%. In the same range, the efficiency of SV1 increases from 60% to 75%. GN1 displays less dependence on  $p_T$  than JetFitter and SV1, efficiently finding upwards of 80% of vertices in  $b$ -jets in this  $p_T$  region. For  $b$ -jets with  $p_T > 100$  GeV, JetFitter finds approximately 65% of vertices, SV1 finds approximately 75% of vertices, and GN1 finds approximately 80% of vertices.

Fig. 6.15 compares the exclusive vertex finding efficiencies of JetFitter and GN1 for multitrack vertices. For  $b$ -jets in the  $Z'$  sample, the vertex finding efficiency drops steeply with increasing  $p_T$  up until  $p_T = 3$  TeV. GN1 outperforms SV1 and JetFitter across the  $p_T$  spectrum. In the first bin, the efficiency of GN1 is 75%, while the efficiencies of SV1 and JetFitter are around 60%. The efficiency of SV1 drops rapidly to almost zero above 3 TeV, while JetFitter and GN1 retain approximately 25% and 30% efficiency respectively. JetFitter finds 45-50% of vertices in  $b$ -jets in the  $t\bar{t}$  sample, while GN1 finds 60-65%. For  $b$ -jets in the  $Z'$  sample, JetFitter finds 35% of vertices in the first bin, dropping to 20% of vertices above 2 TeV. GN1 finds 55% of vertices in the first bin, dropping to 30% above 2 TeV.

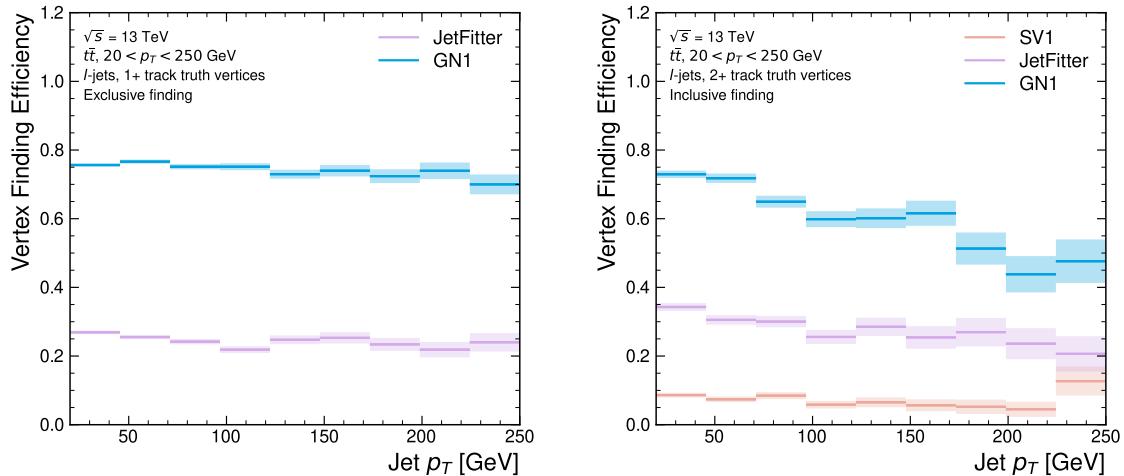


**Figure 6.14:** Heavy flavour vertex finding efficiency as a function of jet  $p_T$  for  $b$ -jets in the  $t\bar{t}$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.



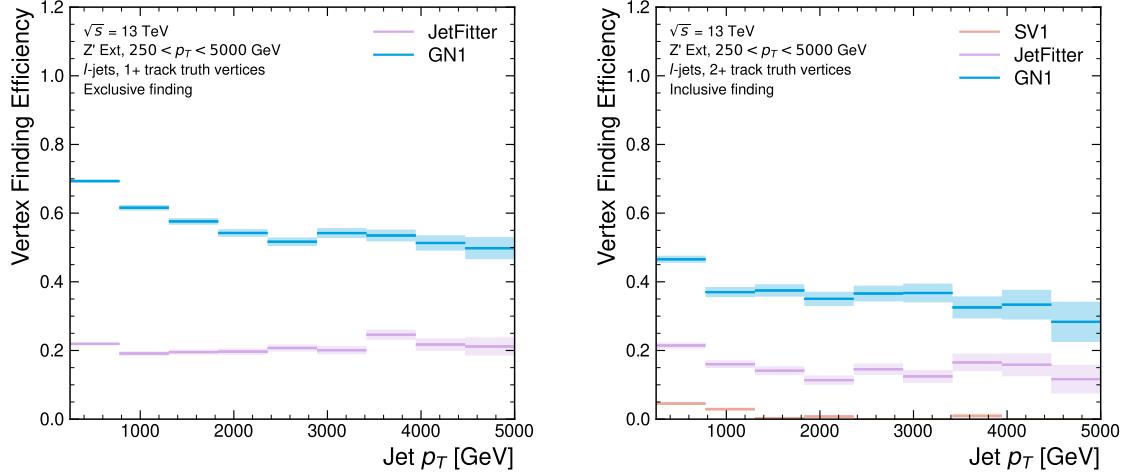
**Figure 6.15:** Heavy flavour vertex finding efficiency as a function of jet  $p_T$  for  $b$ -jets in the  $Z'$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.

1847 While Figs. 6.14 and 6.15 indicate that GN1 is able to successfully find displaced heavy  
 1848 flavour vertices in  $b$ -jets, it is also important to consider the vertexing performance  
 1849 inside light-jets. light-jets may also contain displaced vertices due to long lived  
 1850 secondary particles and material interactions. These tracks have a truth origin of  
 1851 OtherSecondary in the truth labelling scheme enumerated in see Table 5.1. The  
 1852 efficiency to reconstruct vertices comprised of OtherSecondary tracks can be computed  
 1853 in an analogous way to the heavy flavour vertexing efficiency, which is described above.  
 1854 Figs. 6.16 and 6.17 show the efficiency to reconstruct displaced OtherSecondary  
 1855 vertices in light-jets as a function of  $p_T$  for jets in the  $t\bar{t}$  sample and jets in the  $Z'$   
 1856 sample respectively. The figures demonstrate that GN1 is able to more effectively find  
 1857 such vertices in light-jets as compared with SV1 and JetFitter. Since the properties  
 1858 of the displaced vertices in light-jets are likely to be significantly different to heavy  
 1859 flavour vertices found in heavy flavour jets, the improved reconstruction of such  
 1860 vertices may help to differentiate between different flavour of jet.



**Figure 6.16:** Vertex finding efficiency as a function of jet  $p_T$  for light-jets in the  $t\bar{t}$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.

1861 Collectively, the results in this section demonstrate that GN1 is able to accurately  
 1862 group tracks by their spatial origin in both  $b$ -jets and light-jets. The purity of  
 1863 the found vertices was also investigated and was found to be comparable or better  
 1864 than that of SV1 and JetFitter. Using a vertex fitting algorithm to compare the



**Figure 6.17:** Vertex finding efficiency as a function of jet  $p_T$  for light-jets in the  $Z'$  sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions.

reconstructed vertex quantities with those from SV1 and JetFitter is left for future work.

#### 6.4.7 Track Classification Performance

One of the two auxiliary training objectives used by GN1 is to predict the truth origin of each track associated to the jet, as discussed in Section 6.3.2. Since the equivalent information is not provided by any of the existing flavour tagging tools, a benchmark model used to predict the truth origin of each track is trained based on a standard multi-class feed-forward classification network. The benchmark model is trained on the same tracks used for the baseline GN1 training. The model uses precisely the same concatenated track-and-jet inputs as used by GN1 (see Section 6.3.1), but processes only a single track at a time, meaning it cannot take into account the correlations between tracks when determining the track origin. The model is made up of five densely connected linear layers with 200 neurons in each layer. The performance of the model was found to be unsensitive to changes in the network structure.

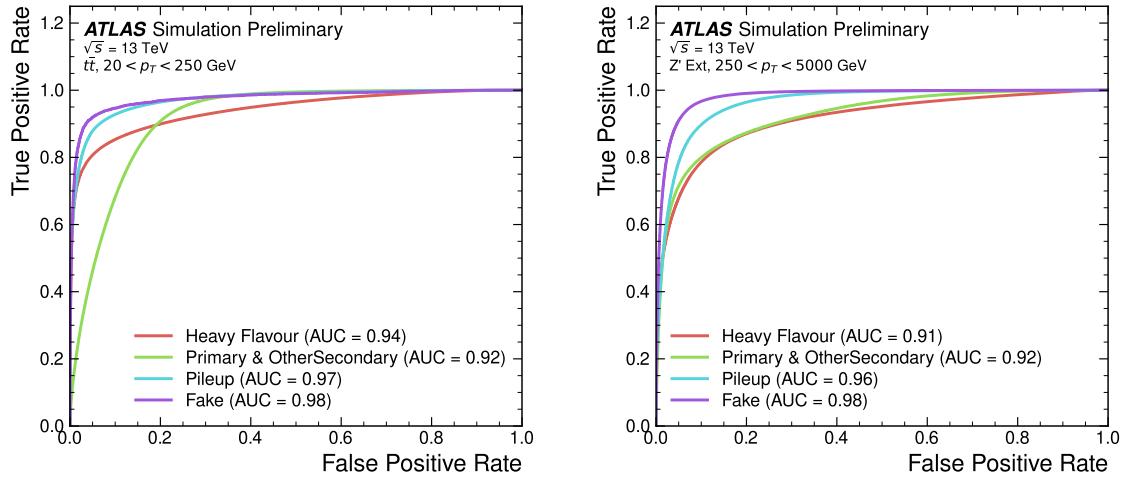
1880 To measure the track classification performance, the area under the curve (AUC)  
 1881 of the receiver operating characteristic (ROC) curve is computed for each origin  
 1882 class, using a one-versus-all classification approach. The AUCs for the different  
 1883 truth origins are averaged using both an unweighted and a weighted mean. The  
 1884 unweighted mean treats the performance of each class equally, while the weighted  
 1885 mean uses as a weight the relative abundance of tracks of each class. Table 6.3  
 1886 demonstrates clearly that GN1 outperforms the MLP both at  $20 < p_T < 250 \text{ GeV}$   
 1887 for jets in the  $t\bar{t}$  sample and at  $250 < p_T < 5000 \text{ GeV}$  for jets in the  $Z'$  sample. For  
 1888 example, GN1 can reject 65% of fake tracks in jets in the  $t\bar{t}$  sample, while retaining  
 1889 more than 99% of good tracks (i.e. those tracks which are not fake). The GN1 model  
 1890 has two advantages over the MLP which can explain the performance improvement.  
 1891 Firstly, the graph neural network architecture enables the sharing of information  
 1892 between tracks as discussed in Section 6.3.3. This is likely to be beneficial since the  
 1893 origins of different tracks within a jet are correlated. Secondly, the jet classification  
 1894 and vertexing objectives may be complementary to the track classification objective,  
 1895 and so the track classification performance is improved by the combined training of  
 1896 complementary objectives.

|            |     | AUC         |             | Precision   |             | Recall      |             | F1          |             |
|------------|-----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|            |     | Mean        | Weighted    | Mean        | Weighted    | Mean        | Weighted    | Mean        | Weighted    |
| $t\bar{t}$ | MLP | 0.87        | 0.89        | 0.39        | 0.71        | 0.51        | 0.56        | 0.36        | 0.62        |
|            | GN1 | <b>0.92</b> | <b>0.95</b> | <b>0.51</b> | <b>0.82</b> | <b>0.64</b> | <b>0.70</b> | <b>0.51</b> | <b>0.74</b> |
| $Z'$       | MLP | 0.90        | 0.94        | 0.36        | 0.84        | 0.47        | 0.72        | 0.31        | 0.76        |
|            | GN1 | <b>0.94</b> | <b>0.96</b> | <b>0.48</b> | <b>0.88</b> | <b>0.60</b> | <b>0.79</b> | <b>0.48</b> | <b>0.82</b> |

**Table 6.3:** The area under the ROC curves (AUC) for the track classification from GN1, compared to a standard multilayer perceptron (MLP) trained on a per-track basis. The unweighted mean AUC over the origin classes and weighted mean AUC (using as a weight the fraction of tracks from the given origin) is provided. GN1, which uses an architecture that allows track origins to be classified in a conditional manner as discussed in Section 6.3.3, outperforms the MLP model for both  $t\bar{t}$  and  $Z'$  jets.

1897 Fig. 6.18 shows the track origin classification ROC curves for the different track  
 1898 origins for jets in both the  $t\bar{t}$  and  $Z'$  samples. In order to improve visual readability  
 1899 of the plot, the curves for the heavy flavour truth origins (FromB, FromBC and  
 1900 FromC) have been combined (weighted by their relative abundance), as have the  
 1901 Primary and OtherSecondary origins. In jets in both the  $t\bar{t}$  and  $Z'$  samples, the AUC

of all the different origin groups exceeds 0.9, representing strong overall classification performance. In both samples fake tracks are the easiest to classify, followed by pileup tracks. The FromC tracks which are  $c$ -hadron decay products, are the hardest to classify, possibly due to their similarity to both fragmentation tracks and  $b$ -hadron decay tracks, depending on the  $c$ -hadron species in question.



**Figure 6.18:** ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the  $t\bar{t}$  sample (left) and jets in the  $Z'$  sample [2]. The FromB, FromBC and FromC labels have been combined, weighted by their relative abundance, into the Heavy Flavour category, and the Primary and OtherSecondary labels have similarly been combined into a single category. The mean weighted area under the ROC curves (AUC) is similar for both samples.

#### 6.4.8 Looser Track Selection

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

$$N_{\text{shared}}^{\text{Si}} = \frac{N_{\text{shared}}^{\text{Pix}} + N_{\text{shared}}^{\text{SCT}}}{2} \quad (6.7)$$

where  $N_{\text{shared}}^{\text{Pix}}$  is the number of shared pixel hits and  $N_{\text{shared}}^{\text{SCT}}$  is the number of shared SCT modules on a track. The nominal cut used elsewhere in this thesis is  $N_{\text{shared}}^{\text{Si}} < 2$ . As the rate of shared hits is significantly higher for  $b$ -hadron decay tracks than for

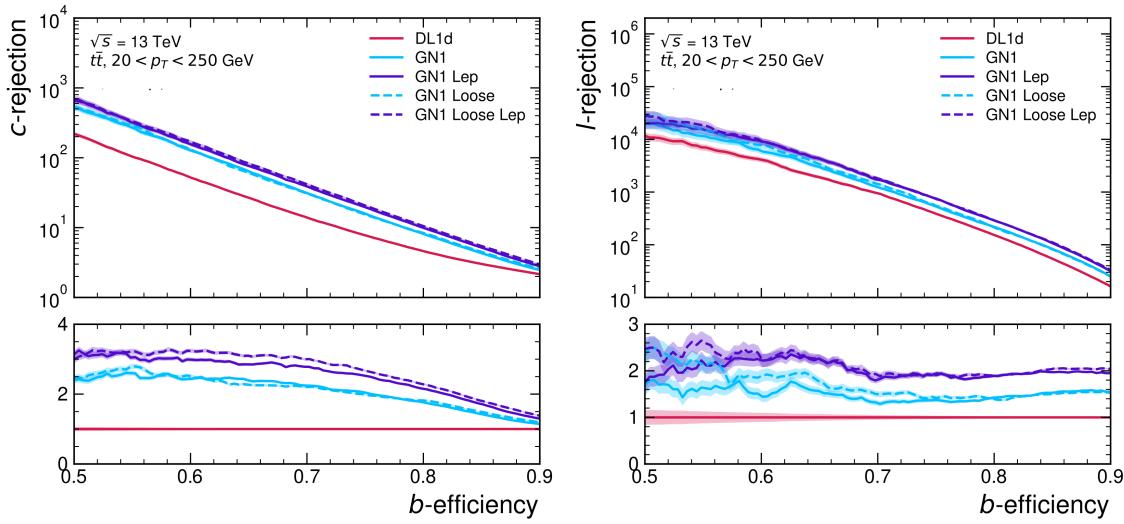
1914 other tracks, especially at high- $p_T$ , this cut rejects a significant proportion of these  
1915 tracks.

1916 Figs. 6.19 and 6.20 show the result of training the GN1 tagger with the full relaxation  
1917 of this cut, i.e. allowing tracks with any number of shared hits. The shared hit  
1918 requirements applied by the ambiguity solver as part of track reconstruction (see  
1919 Section 3.4.1) are still applied. In addition, the maximum allowed value of  $d_0$  is  
1920 increased from 3.5 mm to 5.0 mm. The results show that optimisation of the input  
1921 track selection can lead to significant improvements in performance over the default  
1922 selection. For the jets in the  $t\bar{t}$  sample shown in Fig. 6.19, the effect of loosening  
1923 the track selection is limited. This is expected due to the lower prevalence of shared  
1924 hits at highly displaced tracks at lower transverse momenta. However for jets in the  
1925  $Z'$  sample as shown in Fig. 6.20, the light-jet rejection improves with respect to the  
1926 baseline GN1 model by 30%, while the light-jet rejection improves by 70% at the  
1927 50%  $b$ -jet WP.

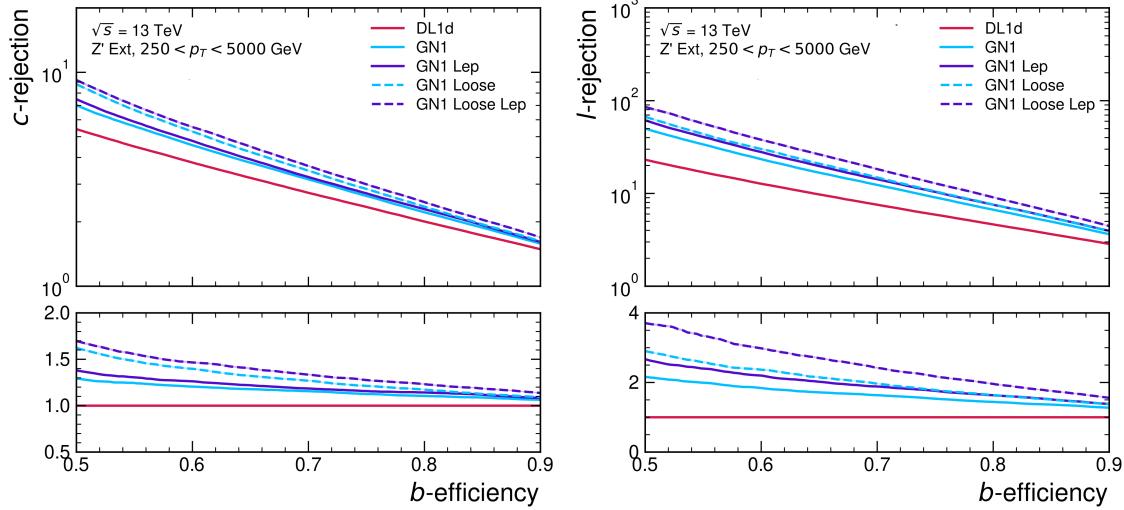
1928 Although the results demonstrate a significant performance improvement at high- $p_T$ ,  
1929 it is also possible that additional studies on further loosening the selection could yield  
1930 further improved results. For example the selections on the number of number of  
1931 holes and the longitudinal impact parameter could be further relaxed. The maximum  
1932 number of tracks provided as input to the model could also be increased from the  
1933 default value of 40. In order to change the default tracking setup, studies investigating  
1934 the modelling uncertainties of the additional tracks need to be carried out.

## 1935 6.5 Other Implementations of GN1

1936 The implementation of GN1 described in this chapter has been re-used in several  
1937 other contexts, demonstrating its flexibility to easily provide good jet flavour tagging  
1938 performance with minimal overhead. The model has been implemented as a  $b$ -jet  
1939 tagger in the High Level Trigger (HLT) (see Section 3.3.4). The inputs to the  
1940 model are the running on precision tracks and jet level quantities reconstructed after  
1941 primary vertexing. Fig. 6.21 shows the performance of GN1 versus a comparable  
1942 DL1d model [124], and two versions of DIPS [81], with EMTopo and PFlow jets (see  
1943 Section 3.4.3) based on a low-precision region-of-interest based tracking pass. The

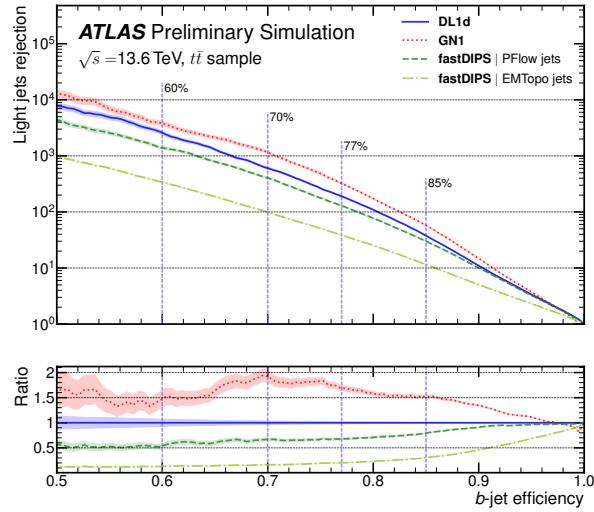


**Figure 6.19:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $t\bar{t}$  jets with  $20 < p_T < 250$  GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1d and  $f_c = 0.05$  is used for GN1. Binomial error bands are denoted by the shaded regions.



**Figure 6.20:** The  $c$ -jet (left) and light-jet (right) rejections as a function of the  $b$ -jet tagging efficiency for  $Z'$  jets with  $250 < p_T < 5000$  GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. A value of  $f_c = 0.018$  is used in the calculation of  $D_b$  for DL1d and  $f_c = 0.05$  is used for GN1. Binomial error bands are denoted by the shaded regions.

1944 trigger implementation of GN1 improves upon the light-jet rejection of DL1d by 50%  
1945 at the 60%  $b$ -jet WP for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV.

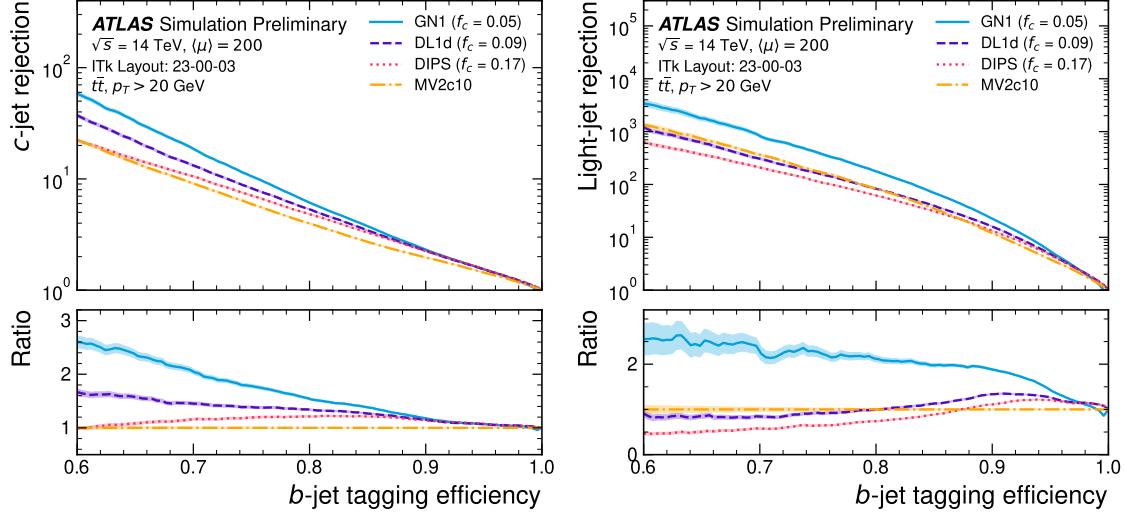


**Figure 6.21:** The light-jet rejection as a function of the  $b$ -jet efficiency jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV for events with a centre of mass energy  $\sqrt{s} = 13.6$  TeV [125]. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for  $b$ -tagging.

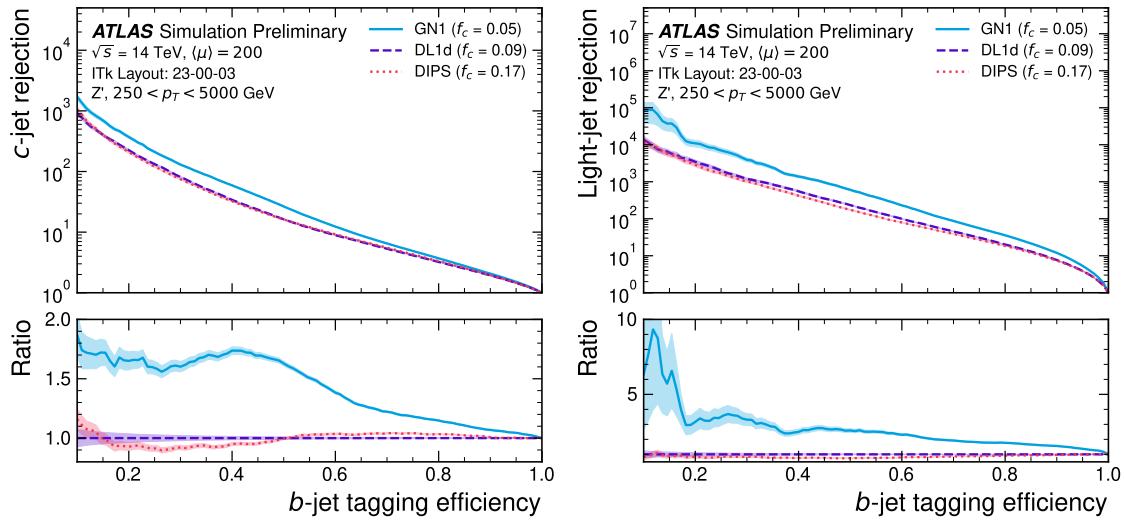
1946 The model also demonstrates strong performance for the High Luminosity LHC  
1947 (HL-LHC), as documented in Ref. [126]. Figs. 6.22 and 6.23 are reproduced from  
1948 Ref. [126]. The results show that GN1 outperforms other existing flavour tagging  
1949 algorithms when trained on an entirely different detector geometry. When compared  
1950 with DL1d [124], GN1 improves on the  $c$ -jet rejection (light-jet rejection) by a factor  
1951 of  $\sim 2$  ( $\sim 2.5$ ) for jets in the  $t\bar{t}$  sample at the 60%  $b$ -jet WP. Significant improvements  
1952 in rejections are also observed for jets in the  $Z'$  sample.

## 6.6 Conclusion

1953 In this chapter a novel jet tagger, GN1, is presented. The model has a graph neural  
1954 network architecture and is trained with auxiliary training objectives, which are  
1955 shown to improve the performance of the basic model. GN1 significantly improves



**Figure 6.22:** The  $c$ -jet rejection (left) and light-jet rejection (right) as a function of the  $b$ -jet efficiency for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV for events with a centre of mass energy  $\sqrt{s} = 14$  TeV [126]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for  $b$ -tagging.



**Figure 6.23:** The  $c$ -jet rejection (left) and light-jet rejection (right) as a function of the  $b$ -jet efficiency for jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV for events with a centre of mass energy  $\sqrt{s} = 14$  TeV [126]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions.

flavour tagging performance with respect to DL1r, the current default ATLAS flavour tagging algorithm, when compared in simulated collisions. GN1 improves  $c$ - and light-jet rejection for jets in the  $t\bar{t}$  sample with  $20 < p_T < 250$  GeV by factors of  $\sim 2.1$  and  $\sim 1.8$  respectively at a  $b$ -jet tagging efficiency of 70% when compared with DL1r. For jets in the  $Z'$  sample with  $250 < p_T < 5000$  GeV, GN1 improves the  $c$ -jet rejection by a factor of  $\sim 2.8$  and light-jet rejection by a factor of  $\sim 6$  for a comparative  $b$ -jet efficiency of 30%.

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

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

Further improvements in the  $b$ - and  $c$ -tagging performance are likely possible with a more thorough optimisation of the model architecture, and the integration of additional information from other parts of the ATLAS detector. The addition of other auxiliary training objectives, such as the truth  $b$ -hadron decay radius and transverse momentum, may also yield additional performance gains on top of the gains achieved by loosening the input track selection (demonstrated in Section 6.4.8).

Additional future work includes the verification of the performance of GN1 on collision data, and the full calibration of the model so it can be used by analyses. The flexible nature of the model means it can also be readily applied to other related problems outside of standard  $b$ - and  $c$ -tagging applications, as demonstrated in Section 6.5. Additional applications for the architecture include  $X \rightarrow bb$  and  $X \rightarrow cc$  tagging.

- 1989 The model could also be repurposed as a pileup jet tagger, or general primary and  
1990 secondary vertexing tool.

<sub>1991</sub> **Chapter 7**

<sub>1992</sub> **Boosted VHbb Analysis**

<sub>1993</sub> The Higgs boson, first observed by ATLAS and CMS at the LHC in 2012 [13, 14],  
<sub>1994</sub> is predicted by the standard model to decay primarily to a pair of  $b$ -quarks, with  
<sub>1995</sub> a branching fraction of  $0.582 \pm 0.007$  for  $m_H = 125$  GeV [25]. Observation of this  
<sub>1996</sub> decay mode was reported by ATLAS [112] and CMS [27] in 2018, establishing the  
<sub>1997</sub> first direct evidence for the Yukawa coupling of the Higgs boson to down-type quarks  
<sub>1998</sub> (see Section 2.2.2). The  $H \rightarrow b\bar{b}$  process is also important for constraining the total  
<sub>1999</sub> decay width of the Higgs [127].

<sub>2000</sub> Whilst the dominant Higgs production mechanism at the LHC is gluon-gluon fusion  
<sub>2001</sub> as outlined in Section 2.2.3, this mechanism has an overwhelming QCD multijet  
<sub>2002</sub> background and so overall sensitivity to the Higgs is low. The QCD multijet  
<sub>2003</sub> background refers to events containing one or more strongly produced jets which  
<sub>2004</sub> are not the decay product of heavy resonances, for example  $g \rightarrow q\bar{q}$ . The gluon-  
<sub>2005</sub> gluon fusion channel contains to leading order only jets in the final state, and  
<sub>2006</sub> therefore it is extremely difficult to distinguish signal events from the overwhelming  
<sub>2007</sub> multijet background. The  $H \rightarrow b\bar{b}$  observation therefore searched for Higgs bosons  
<sub>2008</sub> produced in association with a vector boson  $V$  (where  $V$  can be a  $W$  or  $Z$  boson)  
<sub>2009</sub> which subsequently decays leptonically. The leptonic final states allow for leptonic  
<sub>2010</sub> triggering whilst at the same time significantly reducing the multijet background.

<sub>2011</sub> Two full Run 2 dataset analyses were carried out as a follow-up to the  $H \rightarrow$   
<sub>2012</sub>  $b\bar{b}$  observation [112]. Similar to the observation, both measured the associated  
<sub>2013</sub> production of a Higgs with a vector boson, with the Higgs boson decaying to a pair  
<sub>2014</sub> of  $b$ -quarks. The first analysis [128] was focussed on the resolved phase-space, where  
<sub>2015</sub> the Higgs-jet candidate is reconstructed as two distinct jets with radius parameter

2016  $R = 0.4$ . The second analysis [129] was focussed on the boosted phase-space, where  
2017 the Higgs-jet candidate has a sufficiently large transverse momenta such that it can  
2018 be reconstructed as a single jet with a radius parameter of  $R = 1.0$ . This chapter  
2019 will focus on the latter analysis. The analysis is outlined in Section 7.1. Modelling  
2020 studies performed are detailed in Section 7.2, and the results of the analysis are  
2021 presented in Section 7.4. This analysis has been published in Ref. [129]. Figures and  
2022 tables from Ref. [129] are reproduced here.

## 2023 7.1 Analysis Overview

2024 The boosted  $VH$ ,  $H \rightarrow b\bar{b}$  analysis is focused on the high transverse momentum  
2025 regime, which has the benefit of being more sensitive to physics beyond the Standard  
2026 Model [130], but the disadvantage of being more challenging due to the increased  
2027 difficulty in the accurate reconstruction of high transverse momentum physics objects  
2028 (discussed in Chapter 4). In order to focus on the high- $p_T$  regime, the reconstructed  
2029 vector boson  $p_T^V$  is required to be  $p_T^V > 250$  GeV (see Section 7.1.2). Events are  
2030 also split into two  $p_T^V$  bins with the first bin covering  $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$  and  
2031 the second covering  $p_T^V > 400 \text{ GeV}$ , which allows the analysis to benefit from the  
2032 improved signal-to-background in the high- $p_T$  regime.

2033 The previous ATLAS analysis in Ref. [112] was primarily sensitive to vector bosons  
2034 with a more modest  $p_T^V$  boost in the region of 100–300 GeV. In this regime, the  
2035 Higgs candidate was reconstructed using a pair of jets with radius parameter of  
2036  $R = 0.4$ , called small- $R$  jets. However in the high- $p_T$  regime, the decay products  
2037 of the Higgs boson become increasingly collimated and the small- $R$  jets may  
2038 not be individually resolved. In order to enhance the reconstruction of the Higgs  
2039 boson candidate, this analysis uses a large- $R$  jet with radius parameter  $R = 1.0$  to  
2040 reconstruct the Higgs boson candidate (see Section 3.4.3). The Higgs candidate is  
2041 required to have exactly two ghost-assciated (see Section 3.4.3) and  $b$ -tagged variable-  
2042 radius track-jets. The candidate large- $R$  jet is reconstructed using jet substructure  
2043 techniques, in particular it is trimmed by removing soft and wide-angle components,  
2044 which helps to remove particles from the underlying event and pileup collisions [131].  
2045 Refer to Section 3.4.3 for more details on jet reconstruction.

On top of the binning in  $p_T^V$ , selected events are further categorised into the 0-, 1- and 2-lepton channels depending on the number of charged leptons (electrons and muons) present in the reconstructed final state (also referred to as the 0L, 1L, and 2L channels respectively). The 0-lepton channel targets the  $ZH \rightarrow \nu\nu b\bar{b}$  process, the 1-lepton channel targets  $WH \rightarrow \ell\nu b\bar{b}$ , and the 2-lepton channel targets  $ZH \rightarrow \ell\ell b\bar{b}$ , where  $\ell$  is an electron or muon and  $\nu$  is a neutrino. Each channel has a dedicated set of selections which are listed in more detail in Section 7.1.3. Events in the 0- and 1-lepton channels are further split depending on the number of additional small- $R$  jets not matched to the Higgs-jet candidate. The high-purity signal region (HP SR) has zero such jets, while the low-purity signal region (LP SR) has one or more. This is done so that The 0- and 1-lepton channels also make use of a dedicated  $t\bar{t}$  control region for jets with one or more additional  $b$ -tagged small- $R$  jets, described in Section 7.1.4. A complete overview of the different analysis regions is given in Table 7.1.

explain why  
 $t\bar{t}$ bar

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

**Table 7.1:** Summary of the definitions of the different analysis regions [129]. Signal enriched regions are marked with the label SR. There are regions with relatively large signal purity (HP SR) and with low purity (LP SR). Background enriched regions are marked with the label CR. The shorthand “add” stands for additional small- $R$  jets, i.e. number of small- $R$  jets not matched to the Higgs-jet candidate. The medium and high  $p_T^V$  regions are referred to as  $Mp_T^V$  and  $Hp_T^V$ , respectively.

The signal  $VH$ ,  $H \rightarrow b\bar{b}$  yields is extracted from a profile likelihood fit to the large- $R$  jet mass over several signal and control analysis regions, which are described in Sections 7.1.3 and 7.1.4. The diboson background  $VZ$ ,  $Z \rightarrow b\bar{b}$  yield is simultaneously extracted from the fit, and provides a cross check on the signal extraction. The fit model (described henceforth only as “the fit”) is described in more detail in described in Section 7.3.

### 2066 7.1.1 Data & Simulated Samples

2067 Data from centre-of-mass energy  $\sqrt{s} = 13$  TeV proton-proton collisions at the LHC  
 2068 recorded over the course of Run 2 (between 2015 and 2018) were used for the analysis.  
 2069 The resulting dataset corresponds to a total integrated luminosity of  $139 \text{ fb}^{-1}$  (see  
 2070 Fig. 3.4).

2071 An overview of the MC simulated samples used in the analysis is given in Table 7.2.  
 2072 These samples are used to model the signal and background processes relevant to the  
 2073 analysis, with the exception of the multijet background which is modelled using a  
 2074 data-driven technique. Data and simulated events are reconstructed using the same  
 2075 algorithms, and a reweighting is applied to the simulated events in order to match  
 2076 the pile-up distribution observed in the data.

### 2077 7.1.2 Object Reconstruction

2078 The presence of neutrinos in the  $WH \rightarrow \ell\nu b\bar{b}$  and  $ZH \rightarrow \ell\ell b\bar{b}$  signatures can be  
 2079 inferred from a momentum imbalance in the transverse plane Section 3.4.5. The  
 2080 vector boson transverse momentum  $p_T^V$  is reconstructed as the missing transverse  
 2081 energy  $E_T^{\text{miss}}$  in the 0-lepton channel, as the magnitude of the summed  $\mathbf{E}_T^{\text{miss}}$  and  
 2082 charged-lepton momentum in the 1-lepton channel, and as the transverse momentum  
 2083 of the 2-lepton system in the 2-lepton channel (see Section 3.4.5).

2084 Electrons and muons are reconstructed as outlined in Section 3.4.4, and following the  
 2085 approach described in Ref. [112]. Leptons are required to satisfy the selections listed  
 2086 in Table 7.3. Baseline electrons are required to pass the likelihood-based method  
 2087 described in Section 3.4.4, and Signal electron additionally are required to satisfy  
 2088 a tighter likelihood identification selection. Baseline muons are required to pass  
 2089 the ‘loose’ identification described in Ref. [76], while signal muons are required to  
 2090 pass the ‘medium’ identification working point. All signal leptons are required to  
 2091 additionally satisfy a  $p_T > 27$  GeV selection criteria, except for muons in the 1-lepton  
 2092 channel where a cut of 25 GeV is used. The number of baseline leptons is used to  
 2093 categorise the event into the 0-, 1- or 2-lepton channels. The 1- and 2-lepton channels  
 2094 additionally require one signal lepton to be present.

tell tim the  
previou  
snalysis is  
cited here,  
no lepton  
reco info

| Process  | ME generator   | ME PDF                | PS and Hadronisation    | UE model tune | Cross-section order                |
|--|--|-----------------------|-------------------------|---------------|------------------------------------|
| Signal ( $m_H = 125$ GeV and $b\bar{b}$ branching fraction set to 58%) |  |                       |                         |               |                                    |
| $gg \rightarrow WH \rightarrow \ell\nu b\bar{b}$                       | PowHEG-Box v2 [132] + GoSAM [134] + MiNLO [135, 136] | NNPDF3.0NLO (*) [103] | PyTHIA 8.212 [105]      | AZNLO [133]   | NNLO(QCD) + NLO(EW) [137–143]      |
| $qq \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$      | PowHEG-Box v2 + GoSAM + MiNLO                        | NNPDF3.0NLO (*)       | PyTHIA 8.212            | AZNLO         | NNLO(QCD) <sup>(†)</sup> + NLO(EW) |
| $gg \rightarrow ZH \rightarrow \nu\nu b\bar{b}/\ell\ell b\bar{b}$      | PowHEG-Box v2  | NNPDF3.0NLO (*)       | PyTHIA 8.212            | AZNLO         | NLO + NLL [144–148]                |
| Top quark ( $m_t = 172.5$ GeV)   |  |                       |                         |               |                                    |
| $t\bar{t}$   | PowHEG-Box v2 [132, 149]                             | NNPDF3.0NLO           | PyTHIA 8.230            | A14 [106]     | NNLO+NNLL [150]                    |
| s-channel  | PowHEG-Box v2 [132, 151]                             | NNPDF3.0NLO           | PyTHIA 8.230            | A14           | NLO [152]                          |
| t-channel  | PowHEG-Box v2 [132, 151]                             | NNPDF3.0NLO           | PyTHIA 8.230            | A14           | NLO [153]                          |
| $Wt$   | PowHEG-Box v2 [132, 154]                             | NNPDF3.0NLO           | PyTHIA 8.230            | A14           | Approximate NNLO [155]             |
| Vector boson + jets  |  |                       |                         |               |                                    |
| $W \rightarrow \ell\nu$  | SHERPA 2.2.1 [156–159]                               | NNPDF3.0NNLO          | SHERPA 2.2.1 [160, 161] | Default       | NNLO [162]                         |
| $Z/\gamma^* \rightarrow \ell\ell$                                      | SHERPA 2.2.1   | NNPDF3.0NNLO          | SHERPA 2.2.1            | Default       | NNLO                               |
| $Z \rightarrow \nu\nu$   | SHERPA 2.2.1   | NNPDF3.0NNLO          | SHERPA 2.2.1            | Default       | NNLO                               |
| Diboson  |  |                       |                         |               |                                    |
| $qq \rightarrow WW$  | SHERPA 2.2.1   | NNPDF3.0NNLO          | SHERPA 2.2.1            | Default       | NLO                                |
| $qq \rightarrow WZ$  | SHERPA 2.2.1   | NNPDF3.0NNLO          | SHERPA 2.2.1            | Default       | NLO                                |
| $qq \rightarrow ZZ$  | SHERPA 2.2.1   | NNPDF3.0NNLO          | SHERPA 2.2.1            | Default       | NLO                                |
| $gg \rightarrow VV$  | SHERPA 2.2.2   | NNPDF3.0NNLO          | SHERPA 2.2.2            | Default       | NLO                                |

**Table 7.2:** Signal and background processes with the corresponding generators used for the nominal samples [129]. If not specified, the order of the cross-section calculation refers to the expansion in the strong coupling constant ( $\alpha_s$ ). (\*) The events were generated using the first PDF in the NNPDF3.0NLO set and subsequently reweighted to the PDF4LHC15NLO set [163] using the internal algorithm in POWHEG-BOX v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for the  $pp \rightarrow ZH$  process already includes the  $gg \rightarrow ZH$  contribution. The  $gg \rightarrow ZH$  process is normalised using the cross-section for the  $pp \rightarrow ZH$  process, after subtracting the  $gg \rightarrow ZH$  contribution. An additional scale factor is applied to the  $gg \rightarrow VH$  processes as a function of the transverse momentum of the vector boson, to account for electroweak (EW) corrections at NLO. This makes use of the  $VH$  differential cross-section computed with HAWK [164, 165].

| Variable             | Electrons          | Muons   |
|----------------------|--------------------|---------|
| $p_T$                | $> 7 \text{ GeV}$  |         |
| $ \eta $             | $< 2.47$           | $< 2.7$ |
| $s(d_0)$             | $< 5$              | $< 3$   |
| $ z_0 \sin(\theta) $ | $< 0.5 \text{ mm}$ |         |

**Table 7.3:** Selections applied to baseline and signal electrons and muons.

The analysis makes use of large- $R$  and variable-radius small- $R$  track-jets, which are described in Section 3.4.3. The large- $R$  jets are used to reconstruct the Higgs boson candidate, while the small- $R$  jets are used for  $b$ -tagging and for selection of the analysis region. The track-jets matched to the Higgs candidate are  $b$ -tagged using the MV2c10  $b$ -tagging algorithm (see Chapter 4). The efficiency of the tagging algorithm is calibrated to events in data [166–168]. The jet tagging strategy relies on extensive studies into track-jet  $b$ -tagging in boosted topologies [169, 170].

### 7.1.3 Selection Criteria

An extensive list of selection cuts are applied to each event in order to reject background events whilst retaining as many signal events as possible. A full list of selection cuts applied to the different analysis regions is given in Table 7.4, while some key selections are listed below.

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

discuss  
somewhere,  
but where?

In the 0-lepton channel, trigger selections are applied using an  $E_T^{\text{miss}}$  trigger with a luminosity-dependent threshold between 70–110 GeV. In the 1-lepton electron sub-channel a combination of single electron triggers is used with minimum  $p_T$  thresholds between 24–26 GeV. In the muon sub-channel the same  $E_T^{\text{miss}}$  trigger as the 0-lepton

2118 channel is used. Since muons are not used for the  $E_T^{\text{miss}}$  trigger calculations, this  
2119 is in effect a  $p_T$  requirement on the muon-neutrino system, which in the analysis  
2120 phase space is more efficient than a single-muon trigger. The 2-lepton channel uses  
2121 the same triggering strategy as the 1-lepton channel. In all channels, the trigger  
2122 selections applied are fully efficient for events selected using the full requirements in  
2123 Table 7.4.

2124 The combined selections in Table 7.4 result in a signal efficiency ranging from 6–16%  
2125 for the  $WH$  and  $ZH$  processes depending on the channel and  $p_T^V$  bin.

### 2126 7.1.4 Control Regions

2127 The  $t\bar{t}$  process presents a major background in the 0- and 1-lepton channels. In these  
2128 events, the Higgs candidate is often reconstructed from a correctly tagged  $b$ -jet from  
2129 the top decay  $t \rightarrow Wb$ , and an incorrectly tagged  $c$ - or light-jet from the subsequent  
2130 decay of the  $W$ , as shown in Fig. 7.1.

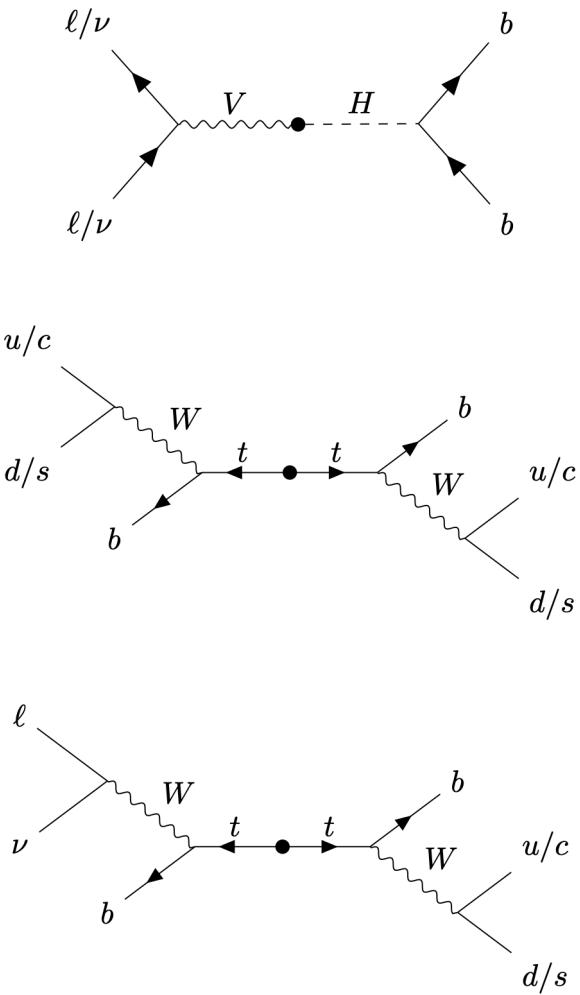
2131 The top quark predominately decays to a  $W$  and a  $b$ -quark. Hence, the second top  
2132 quark from the  $t\bar{t}$  pair is also likely to result in a second tagged  $b$ -tagged track-jet  
2133 outside of the large- $R$  Higgs candidate. To ensure sufficient  $t\bar{t}$  rejection, 0- and  
2134 1-lepton channel signal regions are defined using a veto on events with  $b$ -tagged  
2135 track-jets outside the Higgs-jet candidate. These events are used to construct a  
2136 control region (CR) which is enriched in  $t\bar{t}$  events. The CR is used to constrain the  
2137 normalisation of the  $t\bar{t}$  background in the fit.

### 2138 7.1.5 Background Composition

2139 After the selections described in Section 7.1.3 the number of background events  
2140 mimicking the  $VH$ ,  $H \rightarrow b\bar{b}$  signal is greatly reduced. However, the number of  
2141 background events still greatly outnumbers that of signal events. The background  
2142 processes are channel dependent. In the 0-lepton channel the dominant sources of  
2143 backgrounds are  $Z$ +jets ( $Z \rightarrow \nu\nu$ ) and  $t\bar{t}$ , with  $W$ +jets and diboson events being  
2144 subdominant. In the event of  $W \rightarrow \tau\nu$ , and subsequent hadronic decay of the  $\tau$  or  
2145 lack of successful reconstruction/selection of the leptonic decay products,  $W$ +jets  
2146 can also contribute to the 0-lepton channel.  $t\bar{t}$  and  $W$ +jets (with a leptonic decay

| Selection  | 0 lepton channel    | 1 lepton channel  | 2 leptons channel                      |   |
|--|---------------------|---|--|---|
| Trigger  | $E_T^{\text{miss}}$ | $E_T^{\text{miss}}$   | $e$ sub-channel                        | $\mu$ sub-channel                                 |
| Leptons  | 0 baseline leptons  | 1 signal lepton   | $E_T^{\text{miss}}$                    | Single electron                                   |
|  |                     | $p_T > 27 \text{ GeV}$  | $p_T > 25 \text{ GeV}$                 | $E_T^{\text{miss}}$                               |
|  |                     | no second baseline lepton   |  |   |
|  |                     |   |  | 2 baseline leptons among which                    |
|  |                     |   |  | $\geq 1$ signal lepton, $p_T > 27 \text{ GeV}$    |
|  |                     |   |  | both leptons of the same flavour                  |
|  |                     |   |  | -   |
|  |                     |   |  | opposite sign muons                               |
| $E_T^{\text{miss}}$  | $> 250 \text{ GeV}$ | $> 50 \text{ GeV}$  | -                                      | -   |
| $p_T^V$  |                     |   | $p_T^V > 250 \text{ GeV}$              | -   |
| Large- $R$ jets  |                     | at least one large- $R$ jet, $p_T > 250 \text{ GeV},  \eta  < 2.0$                                    |  |   |
| Track-jets   |                     | at least two track-jets, $p_T > 10 \text{ GeV},  \eta  < 2.5$ , matched to the leading large- $R$ jet |  |   |
| $b$ -tagged jets   |                     | leading two track-jets matched to the leading large- $R$ must be $b$ -tagged (MV2c10, 70%)            |  |   |
| $m_J$  |                     |   | $> 50 \text{ GeV}$                     |   |
| $\min[\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{small-}R \text{ jets})]$        | $> 30^\circ$        |   | -                                      |   |
| $\Delta\phi(\mathbf{E}_T^{\text{miss}}, H_{\text{cand}})$                          | $> 120^\circ$       |   | -                                      |   |
| $\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T, \text{trk}}^{\text{miss}})$ | $< 90^\circ$        |   | -                                      |   |
| $\Delta y(V, H_{\text{cand}})$   | -                   |   | $ \Delta y(V, H_{\text{cand}})  < 1.4$ |   |
| $m_{\ell\ell}$   | -                   |   |  | $66 \text{ GeV} < m_{\ell\ell} < 116 \text{ GeV}$ |
| Lepton $p_T$ imbalance   | -                   |   |  | $(p_T^{\ell_1} - p_T^{\ell_2})/p_T^Z < 0.8$       |

**Table 7.4:** Event selection requirements for the boosted  $VH$ ,  $H \rightarrow b\bar{b}$  analysis channels and sub-channels [129]. Various channel dependent selections are used to maximise the signal efficiency and reduce the number of background events in each analysis region. The  $\min[\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{small-}R \text{ jets})]$  selection is used to remove jets when the missing transverse momentum  $\mathbf{E}_T^{\text{miss}}$  is pointing in the direction of the Higgs candidate, and the  $\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T, \text{trk}}^{\text{miss}})$  is used to reject events where the calorimeter missing transverse momentum  $\mathbf{E}_T^{\text{miss}}$  is not pointing in the direction of the track-based missing transverse momentum  $\mathbf{E}_{T, \text{trk}}^{\text{miss}}$ . The  $\Delta y(V, H_{\text{cand}})$  quantifies the rapidity difference between the reconstructed vector boson and Higgs candidate.



**Figure 7.1:** Diagrams of the signal process (top) and examples of the the 0-lepton and 1-lepton  $t\bar{t}$  backgrounds (middle, bottom). Object to the right of centre are reconstructed within the large- $R$  jet. For the backgrounds, the large- $R$  jet contains a mis-tagged  $c$ - or light-jet. Not shown is the 0-lepton background resulting from  $W \rightarrow \ell\nu$ , where the lepton is out of acceptance or not reconstructed.

2147 of the  $W$  as in  $W \rightarrow \ell\nu$ ) are dominant in the 1-lepton channel, while single-top is  
2148 subdominant. In the 2-lepton channel,  $Z + \text{jets}$  ( $Z \rightarrow \ell\ell$ ) is again dominant followed  
2149 by  $Z Z$  diboson events.

2150 The diboson background  $VV$  consists primarily of  $W Z$  and  $Z Z$  events in which the  
2151  $Z$  decays to a pair of  $b$ -quarks. This process very closely matches the signal, with  
2152 a resonant peak occurring at  $m_Z = 91$  GeV and so is considered as an irreducible  
2153 background ( $V + b$ -jets is also irreducible).

2154 The  $t\bar{t}V$ ,  $t\bar{t}H$  and multijet backgrounds are negligible in the analysis phase space  
2155 after the selections have been applied, with the exception of the 1-lepton electron  
2156 sub-channel, in which multijet background is not negligible. The multijet background  
2157 is made up of events where the isolated leptonic signature has been mimicked by  
2158 either a jet or a muon or electron from a semi-leptonic heavy flavour decay, where  
2159 the lepton has escaped the jet.

2160 The contributions from the different backgrounds are modelled using Monte Carlo  
2161 event generators and the uncertainties associated with these samples are studied in  
2162 Section 7.2. The multijet background is modelled using a data-driven technique.

## 2163 7.2 Systematic Uncertainties & Background 2164 Modelling

2165 Systemic uncertainties are extensively employed to give the fit model enough flexibility  
2166 to account for inaccuracies in the various inputs. Two main types of systematic  
2167 uncertainty are considered: experimental and modelling. Experimental uncertainties  
2168 arise due to the limited precision of the detector, imperfect reconstruction algorithms  
2169 (in particular the  $b$ -tagging algorithms), and due to the imperfect measurement of pile-  
2170 up and integrated luminosity. Modelling uncertainties arise due to the imperfections  
2171 in the Monte-Carlo simulations used to model the signal and background events. In  
2172 order to observe a certain process, for example  $VH$ ,  $H \rightarrow b\bar{b}$ , an increase in the  
2173 number of observed events with respect to the background-only hypothesis is looked  
2174 for. The excess is often relatively small against the total number of background  
2175 events, and hence accurate modelling of the expected number of background and

2176 signal events is crucial for successfully performing the analysis. Particular care is  
2177 paid to the uncertainties on the modelling predictions as discussed in this section.

2178 Modelling uncertainties are described in detail in the following sections. *Nominal*  
2179 samples are used as a reference to which different variations can be compared.  
2180 The nominal samples are chosen as the best possible representation of the underlying  
2181 physical process. *Alternative* samples are used to understand inaccuracies that may  
2182 be present in the nominal samples. Some aspect of the nominal model is varied, and  
2183 the discrepancy with respect to the nominal model is quantified. The discrepancy is  
2184 used to estimate a systematic uncertainty associated with the model parameter which  
2185 was varied. The alternative samples are sometimes obtained via internal weight  
2186 variations or parameterisation methods, rather than by re-running the simulation.  
2187 This is discussed in more detail in Section 7.2.1.

2188 Modelling studies involving  $c$ - and light-jets is hampered by the low number of events  
2189 available after the analysis selection is applied, due to the high rejection rates of the  
2190  $b$ -tagging algorithm MV2c10. For modelling studies, truth tagging (TT) is therefore  
2191 employed to ensure sufficient numbers of jets are available to calculate uncertainties.  
2192 TT works by computing a 2-dimensional efficiency map using the jet  $p_T$  and  $\eta$ . The  
2193 two leading track-jets associated to the large- $R$  jet are weighted based on their  $p_T$   
2194 and  $\eta$  using the pre-calculated efficiency map, rather than being required to explicitly  
2195 pass the  $b$ -tagging requirement.

### 2196 7.2.1 Implementation of Variations

2197 Modelling variations are implemented in different ways, depending on the associ-  
2198 ated uncertainty. Table 7.5 lists the different sources of uncertainty described in  
2199 Section 7.2.2 and for each lists the implementation. As production of high-stastic

2200 MC samples is computationally expensive, a technique in state of the art simulation  
2201 packages is to store some sources of variation as internal weights, which can be  
2202 generated alongside the nominal samples, saving computation time. The nominal  
2203 sample then effectively contains information about an ensemble of different samples,  
2204 corresponding to different model parameters, which are accessible via reweightings.  
2205 When filling histograms for the variations, bins are incremented by the internal  
2206 weight of the event associated with the variation in question.

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

**Table 7.5:** Different sources of uncertainty (i.e. variations in the model) considered for the  $V+jets$  background, and the corresponding implementation. For each uncertainty, acceptance and shape uncertainties are derived.

2207 While the inclusion of internal weight variation in MC event generators has decreased  
 2208 simulation times and increased available statistics, there are in SHERPA 2.2.1 currently  
 2209 some sources of systematic uncertainty that are unable to be stored as internal weight  
 2210 variations due to technical limitations. Two examples are the choice of resummation  
 2211 and merging scales. A method to parameterise the systematic variation using  
 2212 one sample, and to then apply this parameterisation to another sample, has been  
 2213 developed by ATLAS [171]. This method was used to derive resummation and  
 2214 merging uncertainties for the nominal SHERPA 2.2.1 sample, using a previous (lower  
 2215 statistic) SHERPA 2.1 alternative sample. The resulting uncertainties were studied  
 2216 and found to be negligible in comparison with systematics from other sources.

## 2217 7.2.2 Sources of Systematic Uncertainties

2218 This section briefly describes the different sources of uncertainty in the analysis, and  
 2219 how each is implemented. For each source of uncertainty, acceptance and shape  
 2220 uncertainties are derived. Acceptance uncertainties account for the uncertainty in  
 2221 the overall number of events in each channel, and for the migration of events between  
 2222 different analysis regions. Meanwhile, shape uncertainties account for differences in  
 2223 the shapes, but not overall normalisations, of the large- $R$  jet mass.

**2224 QCD Scales**

2225 The  $V + \text{jets}$  matrix element calculations contains infrared and ultraviolet divergences.  
2226 These are handled by introducing arbitrary parameters corresponding to the renormalisation scale ( $\mu_R$ ) and factorisation scale ( $\mu_F$ ). Physical observables are not  
2227 dependent on these parameters when using the infinite perturbation series expansion,  
2228 however at some fixed order in QCD a limited dependence is present. To assess the  
2229 impact of this, both  $\mu_R$  and  $\mu_F$  are independently varied from their nominal values  
2230 by factors of 0.5 and 2 to account for higher order corrections to the calculation of  
2231 the matrix element used to simulate the process.

**2233 PDF Sets**

2234 Parton distribution functions (PDFs) specify the probability of finding a parton with  
2235 a given momentum inside a hadron (in this case, colliding protons). PDFs have  
2236 to be derived from data and are a significant source of uncertainty in analyses of  
2237 hadronic collision data. There are three sources of PDF uncertainties: the statistical  
2238 and systematic errors on the underlying data used to derive the PDFs, the theory  
2239 which is used to describe them (which is based on some fixed order perturbative  
2240 QCD expansion), and finally the procedure which is used to extract the PDFs from  
2241 the data. PDF-related uncertainties were derived following Ref. [163]. This involves  
2242 considering 100 PDF replicas which, when combined, form a central value and  
2243 associated uncertainty, and also in parallel direct changes to the central values of  
2244 PDFs using the MMHT2014 [172] and CT14NLO [173] PDF sets.

**2245 Event Generator**

2246 The choice of parton shower (PS) and underlying event (UE) generators can affect  
2247 the analysis outcome. Changing these models modifies several aspects of the event  
2248 generation at the same time, such as the accuracy of matrix element predictions and  
2249 different approaches to parton showering. This change tends to lead to the largest  
2250 discrepancy with respect to the nominal samples.

---

## 2251 Resummation and Merging Scales

2252 Resummation is a technique used in QCD to help cope with calculations involving  
 2253 disparate energy scales, and involves the introduction of an associated resummation  
 2254 scale, the choice of which introduces some systematic uncertainty into the model.  
 2255 Parton showering models are accurate when simulating low- $p_T$  radiation, however  
 2256 inaccuracies start to arrive when simulating hard emissions. To combat this, parton  
 2257 showering models utilise more precise matrix element calculations above some  
 2258 momentum threshold. The choice of threshold, or *merging scale* introduces some  
 2259 uncertainty into the final result. Resummation (QSF) and merging (CKKW) scale  
 2260 variations are available for a subset of the SHERPA samples. The number of available  
 2261 events is significantly lower than the number of events in the nominal sample, and no  
 2262 statistically significant discrepancy with respect to the nominal samples is observed.  
 2263 The corresponding uncertainties and therefore neglected.

## 2264 7.2.3 Vector Boson + Jets Modelling

2265 After event selection, the  $V+jets$  background is a dominant background in all three  
 2266 analysis channels as described in Section 7.1.5. The  $V+jets$  samples are split into  
 2267 categories depending on the truth flavour of the track-jets which are ghost-associated  
 2268 to the large- $R$  jet Higgs candidate. The categories are  $V+bb$ ,  $V+bc$ ,  $V+bl$ ,  $V+cc$ ,  
 2269  $V+cl$ ,  $V+ll$ , and  $V+hf$  refers collectively to the categories containing at least one  $b$ -  
 2270 or  $c$ -jet.  $V+bb$  is dominant generally accounting for 80% of the events, while  $V+hf$   
 2271 accounts for around 90% of the events. The full flavour composition breakdown for  
 2272 each channel and analysis region are given in Tables 7.6, 7.8 and 7.9.

2273 In order to access uncertainties associated with the use of MC generators, variations  
 2274 of the simulation are produced using alternative generators or variation of nominal  
 2275 generator parameters as described in Section 7.2.1. As described in Section 7.1.1,  
 2276 the nominal MC event generator used for  $V+jets$  events is SHERPA 2.2.1, while  
 2277 MADGRAPH5\_AMC@NLO+PYTHIA8 (which uses a different parton showering  
 2278 model) is used as an alternative generator.

2279 Modelling systematics can have several impacts, including affecting the overall  
 2280 normalisation for different processes, the relative acceptances between different  
 2281 analysis regions (i.e. migrations between HP and LP SRs, between the SR and CR,

and between  $p_T^V$  bins), and the shapes of the  $m_J$  distributions. Since the fit model fits only the large- $R$  jet mass  $m_J$  to data, all shape uncertainties are estimated with respect to this observable. Several sources of uncertainty, summarised in Section 7.2.2, have been assessed.

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

**Table 7.6:** 0-lepton  $W$ +jets nominal sample flavour composition and total event yield [174].  $Mp_T^V$  refers to the medium  $p_T^V$  region, and  $Hp_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

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

**Table 7.7:** 1-lepton  $W$ +jets nominal sample flavour composition and total event yield [174].  $Mp_T^V$  refers to the medium  $p_T^V$  region, and  $Hp_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

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

**Table 7.8:** 0-lepton  $Z$ +jets nominal sample flavour composition and total event yield [174].  $Mp_T^V$  refers to the medium  $p_T^V$  region, and  $Hp_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

| Channel    | $M p_T^V$         | $H p_T^V$        | $p_T^V$ inclusive |
|------------|-------------------|------------------|-------------------|
| <b>Zbb</b> | 80.80%            | 76.95%           | 79.76%            |
| <b>Zbc</b> | 8.10%             | 6.26%            | 7.60%             |
| <b>Zbl</b> | 4.95%             | 7.06%            | 5.52%             |
| <b>Zcc</b> | 3.97%             | 4.46%            | 4.10%             |
| <b>Zcl</b> | 1.61%             | 3.60%            | 2.14%             |
| <b>Zll</b> | 0.57%             | 1.68%            | 0.87%             |
| Events     | $115.49 \pm 2.42$ | $42.42 \pm 1.27$ | $157.92 \pm 2.73$ |

**Table 7.9:** 2-lepton  $Z + \text{jets}$  nominal sample flavour composition and total event yield [174].  $M p_T^V$  refers to the medium  $p_T^V$  region, and  $H p_T^V$  refers to the high  $p_T^V$  region (see Table 7.1).

## 2286 Acceptance Uncertainties

2287 Several different types of acceptance uncertainties have been calculated and im-  
 2288 plemented as nuisance parameters in the fit. These account for the uncertainty  
 2289 in the overall number of events in each channel, and for the migration of events  
 2290 between different analysis regions. The acceptance uncertainties relevant to the  
 2291  $V + \text{jets}$  processes are summarised below.

- 2292 • **Overall normalisation:** only relevant where normalisation cannot be left  
 2293 unconstrained (or “floating”, i.e. determined as part of the fit). The  $V + \text{hf}$   
 2294 component is left floating in the fit, with independent normalisations used for  
 2295  $W + \text{hf}$  and  $Z + \text{hf}$ . The normalisations are mainly determined by the 1-lepton  
 2296 (for  $W + \text{hf}$ ) and 2-lepton (for  $Z + \text{hf}$ ) regions respectively and then extrapolated  
 2297 to the 0-lepton channel. The negligible  $V + \text{jets}$  backgrounds were constrained to  
 2298 their cross-sections in the fit. were constrained to their cross-sections
- 2299 • **SR-to-CR relative acceptance:** the uncertainty on the relative number of  
 2300  $V + \text{jets}$  events in the signal and control regions.
- 2301 • **HP-to-LP relative acceptance:** the uncertainty on the relative number of  
 2302  $V + \text{jets}$  events in the HP and LP SRs.
- 2303 • **Medium-to-high  $p_T^V$  relative acceptance:** the uncertainty on the relative  
 2304 number of  $V + \text{jets}$  events in the medium and high  $p_T^V$  bins.
- 2305 • **Flavour relative acceptance:** for each flavour  $V + xx$ , where  $xx \in \{bc, bl, cc\}$   
 2306 the ratio of  $V + xx/V + bb$  events is calculated. This corresponds to the uncer-  
 2307 tainty on the heavy flavour composition of the  $V + \text{hf}$  background.

- 2308     • **Channel relative acceptance:** the uncertainty on the relative number of  
 2309        $V+jets$  events between the channels.

2310     The uncertainties arising from the different sources described in Section 7.2.2 are  
 2311       summed in quadrature to give a total uncertainty on each region. A summary of  
 2312       the different acceptance uncertainties that were derived and subsequently applied  
 2313       in the fit are given in Table 7.10. An effort has been made, wherever possible, to  
 2314       harmonise similar uncertainties across different analysis regions and channels.

| V+jets Acceptance Uncertainties |                  |                  |     |    |
|---------------------------------|------------------|------------------|-----|----|
| Boson                           | W                |                  | Z   |    |
| Channel                         | 0L               | 1L               | 0L  | 2L |
| Vbb Norm.                       | 30%              | -                | -   | -  |
| SR-to-CR                        | 90% <sup>†</sup> | 40% <sup>†</sup> | 40% | -  |
| HP-to-LP                        | 18%              |                  | 18% | -  |
| Medium-to-high $p_T^V$          | 30%              | 10%*             | 10% |    |
| Channel relative acceptance.    | 20%              | -                | 16% | -  |
| Vbc/Vbb                         | 30%              |                  |     |    |
| Vbl/Vbb                         | 30%              |                  |     |    |
| Vcc/Vbb                         | 20%              |                  |     |    |
| Vcl Norm.                       | 30%              |                  |     |    |
| Vl Norm.                        | 30%              |                  |     |    |

**Table 7.10:**  $V+jets$  acceptance uncertainties [174].  $W+jets$  SR and CR uncertainties marked with a superscript  $\dagger$  are correlated. The 1L  $W+jets$  H/M uncertainty marked by \* is applied as independent and uncorrelated NPs in both HP and LP signal regions.

### 2315   Shape Uncertainties

2316     In order to derive shape uncertainties for a given background or signal process,  
 2317       normalised distributions of the reconstructed large- $R$  Higgs candidate jet mass  $m_J$   
 2318       are compared for the nominal sample and variations. For each variation, the ratio of  
 2319       the variation to nominal is calculated, the up and down variations are symmetrised,  
 2320       and an analytic function is used to parameterise the ratio. If different analysis regions  
 2321       or channels show the same pattern of variation, a common uncertainty is assigned.

2322 An example of a significant source of uncertainty, arising from choice of factorisation  
 2323 scale  $\mu_R$  is shown in Fig. 7.2. The HP SRs in the medium and high  $p_T^V$  bins are  
 2324 shown for the 0-lepton channel for the  $W+hf$  and  $Z+hf$  jets. The 0- and 1-lepton  
 2325 channels for the  $W+hf$  contribution and the 0- and 2-lepton channels for the  $Z+jets$   
 2326 contribution were found to have compatible shapes in  $m_J$  across channels, and so  
 2327 were jointly measured. An exponential function  $e^{p_0 + p_1 x} + p_2$  has been fitted to the  
 2328 ratio of the normalised distributions. The magnitude of the variation is  $p_T^V$  dependent,  
 2329 and so separate uncertainties are implemented in the fit for each  $p_T^V$  region.

2330 The shape uncertainties for  $\mu_R$  were derived on the SRs but are also applied to the  
 2331 CRs, as the low statistics in the CRs make it difficult to derive dedicated shape  
 2332 uncertainties. All the shape uncertainties are fully correlated accross regions.

remove  
ATLAS  
Simula-  
tion/Internal  
label

2334 A comparison of the  $m_J$  shapes between SHERPA and MADGRAPH is shown in  
 2335 Fig. 7.3. The plots are split by process and channel, but merged in SR purity and  $p_T^V$   
 2336 bins reflecting similarities between the  $m_J$  shapes and variations across these regions.  
 2337 Due to the low statistics available for the alternate MADGRAPH sample, and the  
 2338 lack of statistically significant variation between the samples, no additional shape  
 2339 uncertainty was added to the fit in this case.

remove  
ATLAS  
Simula-  
tion/Internal  
label

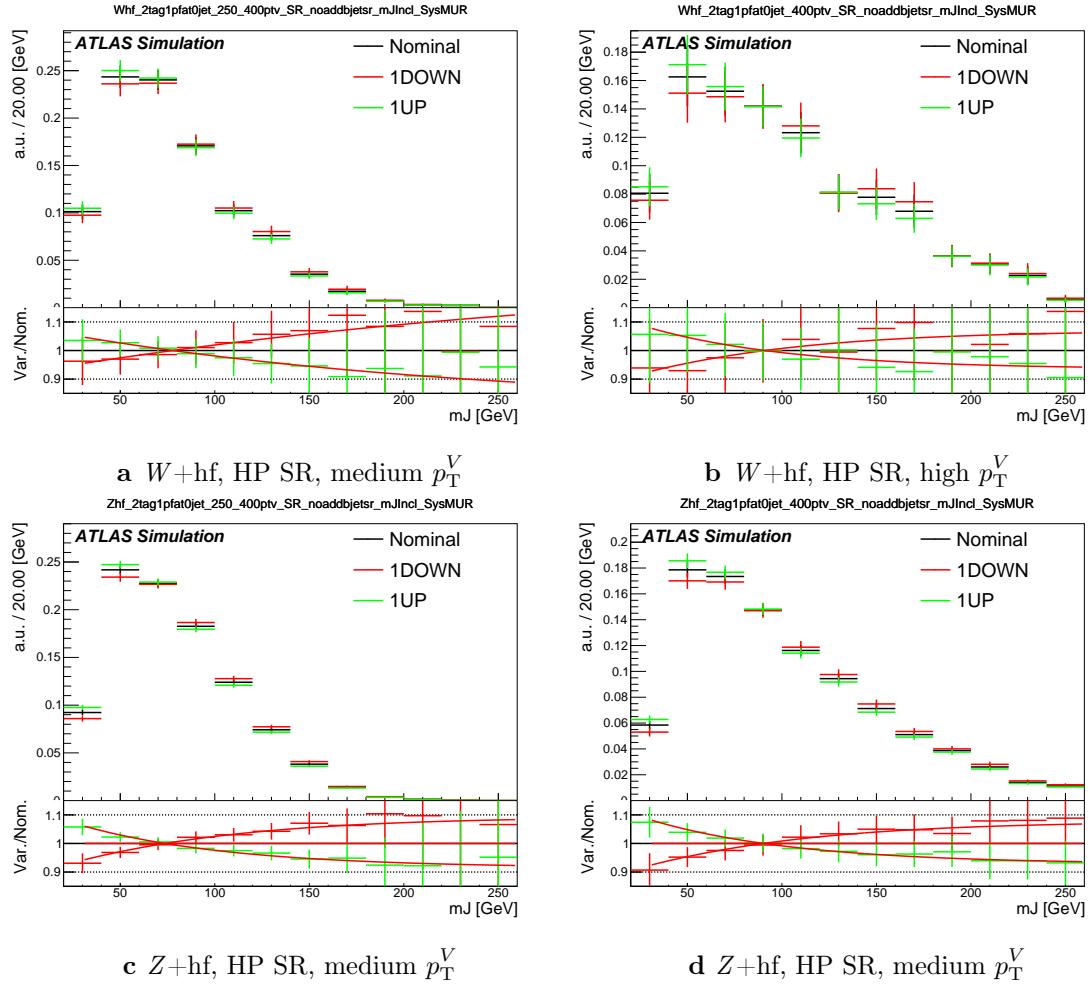
2341 The impacts of variations in the factorisation scale  $\mu_F$  and the choice of PDF set on  
 2342  $m_J$  shape were also found to be negligible in comparison with  $\mu_R$  and are hence no  
 2343 additional uncertainty was added to the fit.

#### 2344 7.2.4 Diboson Modelling

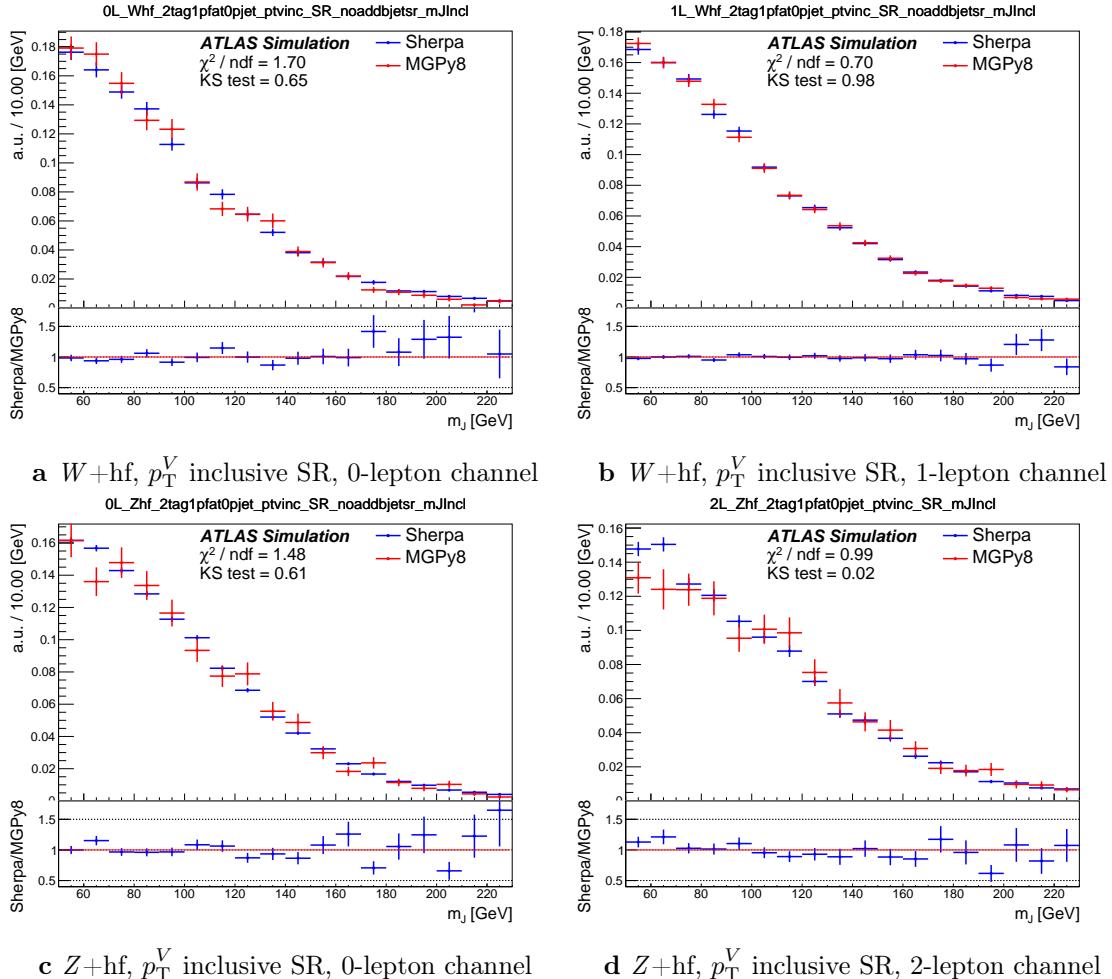
2345 The procedure to derive the uncertainties for the diboson background generally  
 2346 follows that of  $V+jets$ .

2347 The nominal diboson samples are generated using SHERPA 2.2.1 (except for  $gg \rightarrow VV$   
 2348 which uses SHERPA 2.2.2) with the NNPDF3.0NNLO tune. Alternative samples were  
 2349 generated using POWHEG interfaced with PYTHIA8, using the AZNLO shower tune  
 2350 with the CTEQ6L1 PDFs [175]. Unlike SHERPA, POWHEG models the off-shell  $Z$   
 2351 contribution at NLO.

2352 Acceptance and shape uncertainties are derived in an analogous fashion to  $V+jets$   
 2353 as described below.



**Figure 7.2:** Leading large- $R$  jet mass for the  $Z$  and  $W + hf$  processes in the HP SR of the 0-lepton channel [174]. The renormalisation scale  $\mu_r$  has been varied by a factor of 2 (1up) and 0.5 (1down). An exponential function is fitted to the ratio between the nominal and alternate samples.



**Figure 7.3:** Leading large- $R$  jet  $m_J$  inclusive in  $p_T^V$  for the  $V + \text{hf}$  process modelled using both the SHERPA and MADGRAPH samples [174]. The Kolmogorov-Smirnov test and  $\chi^2/\text{ndf}$  are shown on the plots.

2354 **Acceptance Uncertainties**

2355 Diboson acceptance uncertainties are summarised in Table 7.11. Variations from  $\mu_R$ ,  
2356  $\mu_F$ , PDF choice and an alternative generator are considered and are combined via  
2357 a sum in quadrature as described in Section 7.2.3. The largest modification to the  
2358 nominal acceptance results from the POWHEG+PYTHIA8 alternate sample. Since  
2359 the diboson contribution to the  $t\bar{t}$  control region is negligible, no SR-to-CR relative  
2360 acceptance uncertainty is necessary.

2361 For the  $WZ$  contribution, uncertainties are derived using the 1-lepton channel and  
2362 applied to all three channels. An additional 8% channel migration uncertainty  
2363 is applied on the 0-lepton channel. For the  $ZZ$  contribution, the normalisation  
2364 uncertainty is calculated using the 2-lepton channel and applied to all three channels.  
2365 The 0- and 1-lepton channels have a similar HP-to-LP relative acceptance uncertainty  
2366 of 18%. The 1-lepton medium-to-high  $p_T^V$  relative acceptance is based off the value  
2367 obtained from the 2-lepton channel. 30% and 18% channel migration uncertainties  
2368 are applied in the 0- and 1-lepton channels respectively.

2369 Since the contribution from  $WW$  is negligible, dedicated studies are not performed,  
2370 but a 25% normalisation uncertainty is applied in all the three channels which is  
2371 based on the modelling studies performed for the previous analysis [112].

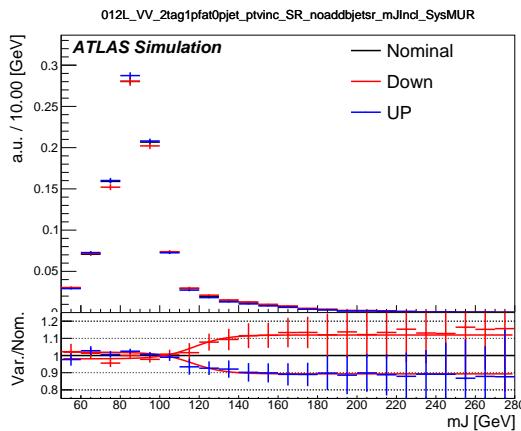
The relevance of this discussion isn't clear or why these choices were made. Needs more detail and discussion

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

**Table 7.11:** Diboson acceptance uncertainties [174]. All uncertainties except channel extrapolation uncertainties are fully correlated between  $ZZ$  and  $WZ$  processes and channels.

## 2372 Shape Uncertainties

2373 Diboson shape uncertainties are derived in a similar fashion to  $V+jets$ . Only the  
 2374 uncertainties associated with systematic variation of  $\mu_R$  and the alternate event  
 2375 generator have a non-negligible impact on the  $m_J$  shape. Variation of  $\mu_R$  produces  
 2376 consistent  $m_J$  shape changes across all regions and channels, and hence only a single  
 2377 associated uncertainty is derived, shown in Fig. 7.4. A hyperbolic tangent is fitted  
 2378 to the symmetrised ratio.



**Figure 7.4:** Leading large- $R$  jet mass distribution for the  $WZ$  and  $ZZ$  processes, inclusive across all signal regions and lepton channels [174]. The renormalisation scale  $\mu_R$  has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function.

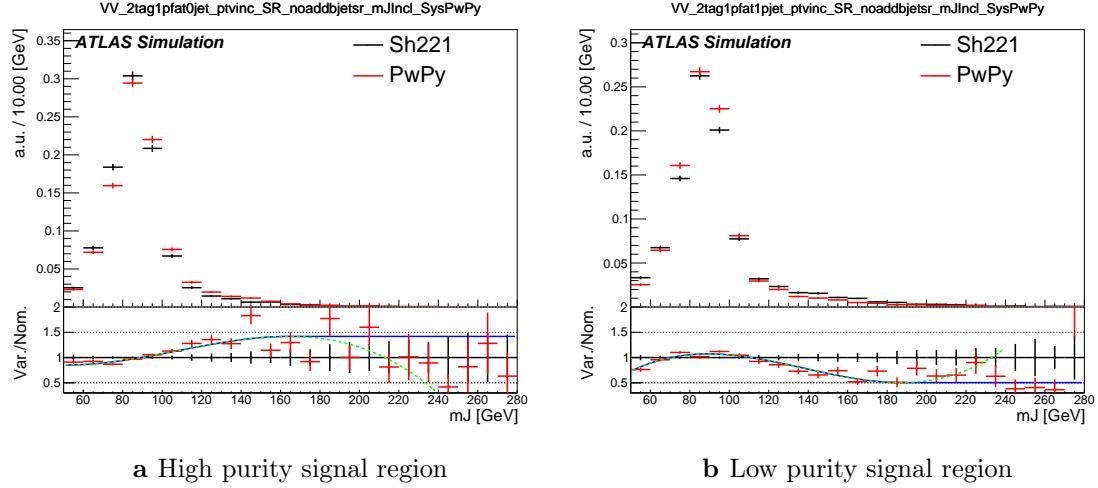
2379

2380 The comparison between the nominal SHERPA and alternative POWHEG+PYTHIA8  
 2381 samples is shown in Fig. 7.5 for the 0- and 1-lepton channels for both the  $WZ$  and  
 2382  $ZZ$  processes. For these channels, the shape of  $m_J$  varies in opposite directions in the  
 2383 LP and HP signal regions. Shapes are similar between  $p_T^V$  bins, the 0- and 1-lepton  
 2384 channels and for  $WZ$  and  $ZZ$ . A third order polynomial is fitted to the ratio, and  
 2385 this function transitions to a constant piecewise function in the high mass region to  
 2386 accurately represent the shape. Dependence on the event generator was found to be  
 2387 negligible within statistical uncertainty in the 2-lepton channel. All diboson shape  
 2388 uncertainties are fully correlated in the fit.

remove  
ATLAS  
Simulation/Internal  
label

More discussion is needed on the relevance of this comments, it isn't clear why this information is stated or what it means in the final implementation. A bit more detail is needed

remove  
ATLAS  
Simulation/Internal  
label



**Figure 7.5:** The comparison on  $m_J$  shapes between SHERPA and POWHEG+PYTHIA 8 samples from  $WZ$  and  $ZZ$  process in high and low purity signal regions, integrated over  $p_T^V$  regions and the 0- and 1-lepton channels [174]. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after a protection is added in the high mass region.

### 2390 7.3 Statistical Treatment

2391 A binned global maximum-profile-likelihood fit of the  $m_J$  distribution is performed  
 2392 to extract information on the signal, combining all the analysis regions defined  
 2393 in Table 7.1. The signal strength  $\mu = \sigma/\sigma_{\text{SM}}$  is defined as the ratio between the  
 2394 observed and predicted cross-sections, where  $\mu = 0$  corresponds to the background-  
 2395 only hypothesis and  $\mu = 1$  corresponds to the SM prediction. This is a parameter of  
 2396 interest (POI) which acts to scale the total number of signal events, and is determined  
 2397 during the fit procedure.

2398 The present analysis makes use of two POIs. The first,  $\mu_{VH}^{bb}$ , is the signal strength  
 2399 for the  $VH$ ,  $H \rightarrow b\bar{b}$  process, the primary process under investigation. The diboson  
 2400 production strength  $\mu_{VZ}^{bb}$  for the  $VZ$ ,  $Z \rightarrow b\bar{b}$  process is measured simultaneously and  
 2401 provides a validation of the analysis apparatus used for the primary  $H \rightarrow b\bar{b}$  measure-  
 2402 ment. Alongside the two POIs, the predictive model depends on several parameters  
 2403 which are not the primary target of measurement, and represent the systematic  
 2404 uncertainties discussed previously. These parameters are called nuisance parameters  
 2405 (NPs), collectively referred to as  $\theta$ . Freely floating background normalisations are  
 2406 implemented as NPs and are also extracted during the fitting processes.

### 2407 7.3.1 Likelihood Function

2408 The statistical setup treats each bin as a Poisson counting experiment and is based on  
 2409 the ROOSTATS framework [176]. The combined likelihood over  $N$  bins is constructed  
 2410 as the product of Poisson probabilities in each bin. Considering the simplified case  
 2411 of a single signal strength parameter  $\mu$ , and neglecting sources of systematic or  
 2412 statistical uncertainty, this is given by

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

2413 where  $s_i$  ( $b_i$ ) is the expected number of signal (background) events in bin  $i$ , and  $n_i$  is  
 2414 the number of observed data events in bin  $i$ .

### 2415 Treatment of Uncertainties

2416 Systematic uncertainties can modify the predicted signal and background yields  $s_i$   
 2417 and  $b_i$ . Each source of systematic uncertainty is taken into account by adding an  
 2418 additional NP  $\theta_j$  to the likelihood in the form of a Gaussian cost function. The  
 2419 combined effect of the NPs is then

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

2420 where  $N_\theta$  is the number of NPs,  $\bar{\theta}_j$  is the nominal value of the  $j$ th NP,  $\theta_j$  is the fitted  
 2421 value, and  $\sigma_j$  is the corresponding associated prior uncertainty on  $\theta_j$ . As the fitted  
 2422 value of the  $\theta_j$  deviates from its nominal value, a cost is introduced. The presence  
 2423 of NPs modifies the likelihood as

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

2424 The predicted signal and background yields are also modified by the presence of the  
 2425 NPs with

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

introduce  
what a pull  
is and what  
a constraint  
is here

- 2426 For NPs which are left freely floating in the fit, no corresponding Gaussian constraints  
2427 are added to the likelihood.
- 2428 The statistical uncertainty on the simulated events is implemented using a dedicated  
2429 NP for each bin which can scale the background yield in that bin. Statistical NPs  
2430 are also implemented using a Gaussian constraint.

### 2431 Smoothing and Pruning

- 2432 To simplify the fit to reduce and improve its robustness, systematic uncertainties  
2433 are smoothed and pruned. Smoothing accounts for the large statistical uncertainty  
2434 present in some samples that can lead to unphysical fluctuations in the shape  
2435 systematics. The smoothing procedure relies on the assumption that the impact of  
2436 systematics should be approximately monotonic and correlated between neighbouring  
2437 bins.
- 2438 In addition to smoothing, pruning is the process of removing from the fit those  
2439 systematics which only have a very small effect. This improves the stability of the  
2440 fit by reducing the number of degrees of freedom. Acceptance uncertainties are  
2441 pruned in a given region if they have a variation of less than 0.5%, or if the up and  
2442 down variations have the same sign in that region. Shape uncertainties are pruned  
2443 in a given region if the deviation in each bin is less than 0.5% in that region. In  
2444 addition, acceptance and shape uncertainties are neglected in a given region for any  
2445 background which makes up less than 2% of the total background in a given region.

### 2446 Fit Procedure and Statistical Tests

- 2447 The best-fit value of  $\mu$ , denoted  $\hat{\mu}$ , is obtained via an unconditional maximisation of  
2448 the likelihood. The likelihood is also used to construct a statistical test which can  
2449 confirm or reject the background-only hypothesis. The test statistic  $q_\mu$  is constructed  
2450 from the profile likelihood ratio,

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

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

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

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

Typically a value of  $Z = 3$  constitutes *evidence* of a processes, while  $Z = 5$  is required for a *discovery*, or observation. Alongside the  $p$ -value, the best-fit value of the signal strength  $\hat{\mu}$  and its corresponding uncertainty are quoted, and compared to their expected values.

### 7.3.2 Background Normalisations

The backgrounds which can be constrained by the fit are left freely floating and the corresponding normalisation factors are extracted. Normalisation factors (NF), represent the value by which the predicted normalisations are scaled, and are implemented for the dominant backgrounds ( $t\bar{t}$ ,  $Z+hf$ ,  $W+hf$ ). The NFs are also subdivided into different regions of phase-space for  $t\bar{t}$ , given it is possible to obtain a strong constraint in the individual channels. This also removes the need for an extrapolation uncertainty.

The normalisations and shapes of all other backgrounds, with the exception of the multijet background which is estimated using a data driven technique, are initialised using the nominal samples and the state-of-the art process normalisations, as outlined in Table 7.2.

### 7.3.3 Asimov Dataset & Expected Results

The Asimov dataset is constructed by replacing the data with the sum of the signal and background predictions  $n_i = s_i + b_i$ . A fit to this dataset using the nominal

2477 values of the NPs from the simulation will recover the input values and is useful for  
2478 studying the expected result, in addition to constraints on and correlations between  
2479 the NPs.

2480 Alternatively, a conditional fit to the Asimov dataset can be performed using values  
2481 of the background NPs which are determined from an unconditional fit to data. The  
2482 signal NPs and POIs are fixed at their nominal values from the SM simulation. The  
2483 result of this fit can be used to calculate expected (median) significances given a  
2484 more realistic background model, which can be compared to their observed values,  
2485 as is done in Section 7.4.2.

## 2486 7.4 Results

2487 In the present analysis, the two signal strength parameters  $\mu_{VH}^{bb}$  and  $\mu_{VZ}^{bb}$  are extracted  
2488 from a simultaneous maximisation of the likelihood described in Section 7.3. The  
2489 results of the analysis are summarised in this section. The corresponding postfit  
2490 background normalisations are listed in Table 7.12. Post-fit  $m_J$  distributions are  
2491 shown in Section 7.4.1. The observed signal strengths are given in Section 7.4.2,  
2492 along with observed and expected significances. Finally in Section 7.4.3 the impact  
2493 of systematic uncertainties on the results is examined.

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

**Table 7.12:** Factors applied to the nominal normalisations of the  $t\bar{t}$ ,  $W+hf$ , and  $Z+hf$  backgrounds, as obtained from the likelihood fit [129]. The errors represent the combined statistical and systematic uncertainties.

### 2494 7.4.1 Post-fit Distributions

2495

Add something in the previous section about how the post-fit plots are produced?

In addition to the observed significance and signal strength, it is also useful to study the post-fit  $m_J$  distributions to compare the agreement between the simulation and data using the best-fit values  $\hat{\mu}$  and  $\hat{\theta}$ . Post-fit  $m_J$  distributions are given for the signal regions in the 0-, 1- and 2-lepton channels in Fig. 7.6. The signal regions are merged for the sake of simplicity. The LP and HP regions are merged for the 0- and 1-lepton channels. In general there is a good level of agreement between the simulation and data, indicating the fit model is performing as expected. Fig. 7.7 shows the post-fit plots for the  $t\bar{t}$  control regions. Again, a good level of agreement is observed given the statistical uncertainties on the distributions.

#### 7.4.2 Observed Signal Strength & Significance

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

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

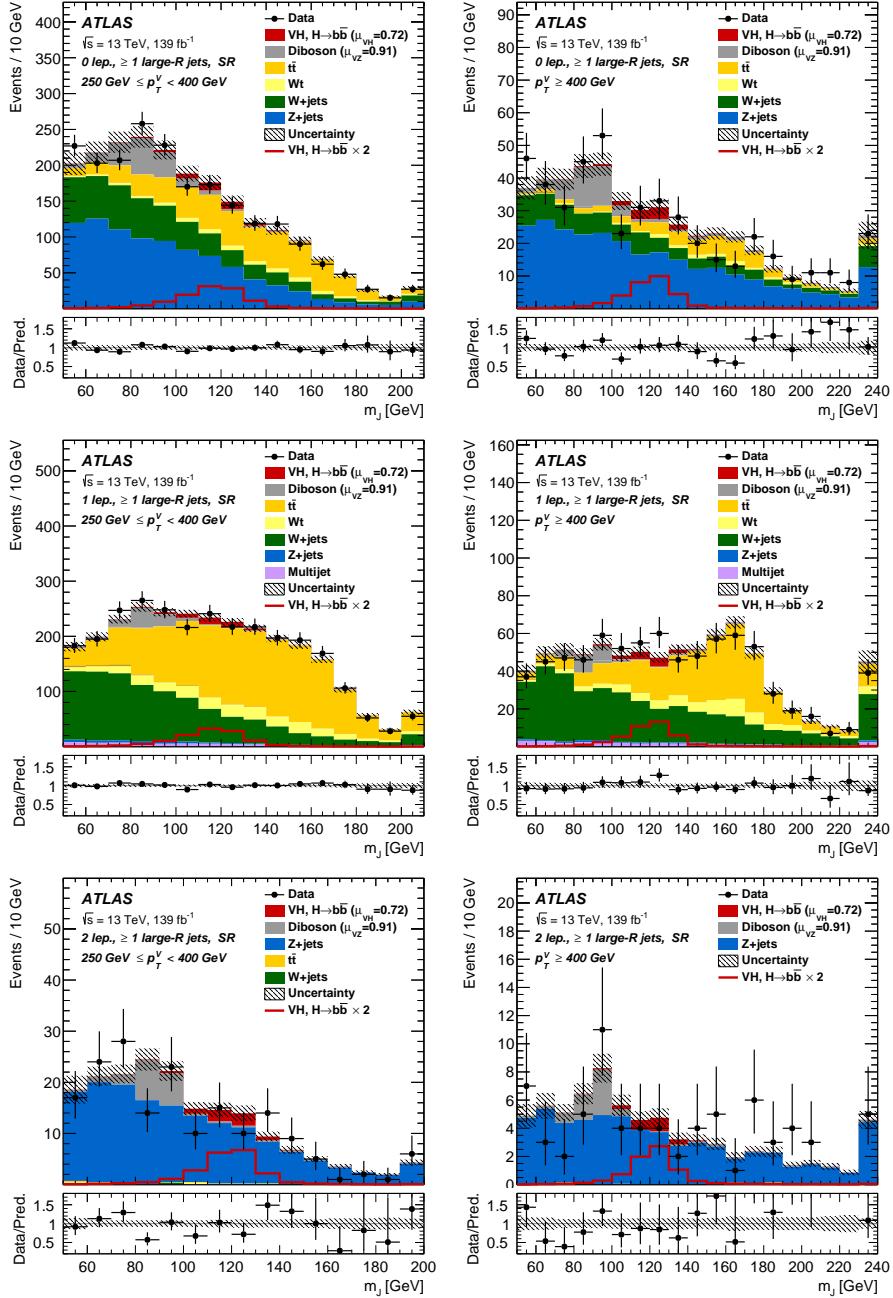
2510

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

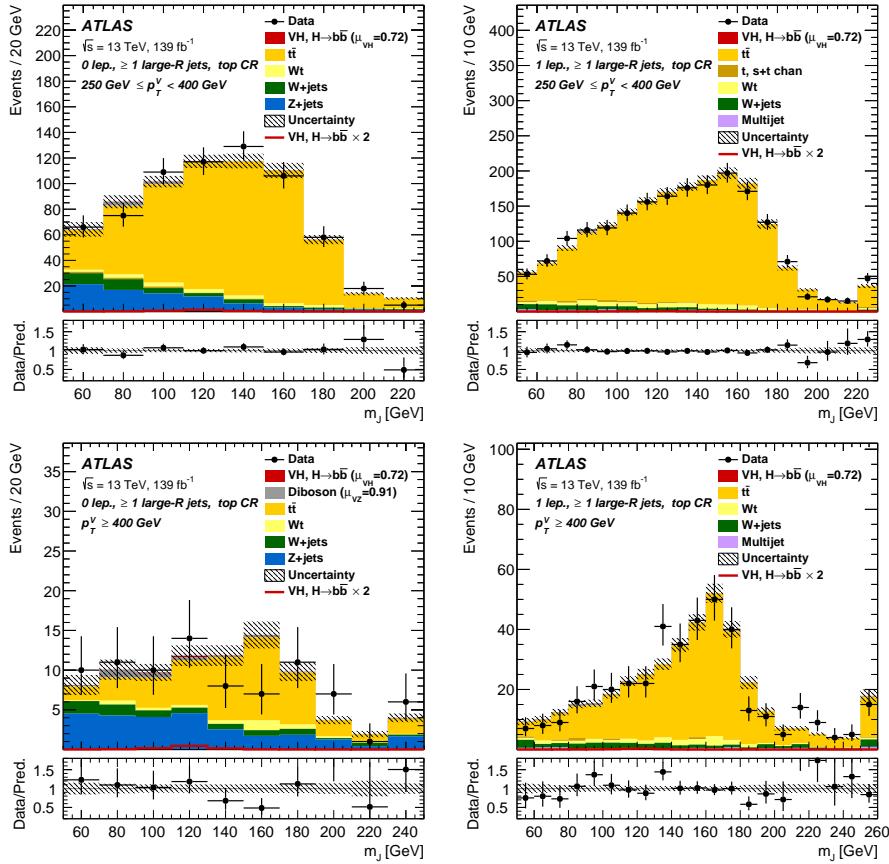
The results for  $\mu_{VH}^{bb}$  and  $\mu_{VZ}^{bb}$  are consistent with the expectation from the SM. The  $\mu_{VH}^{bb}$  measurement is dominated by statistical uncertainty, while the  $\mu_{VZ}^{bb}$  measurement is dominated by systematic sources of uncertainty. The measured signal strength for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

#### 2519 Compatability Studies

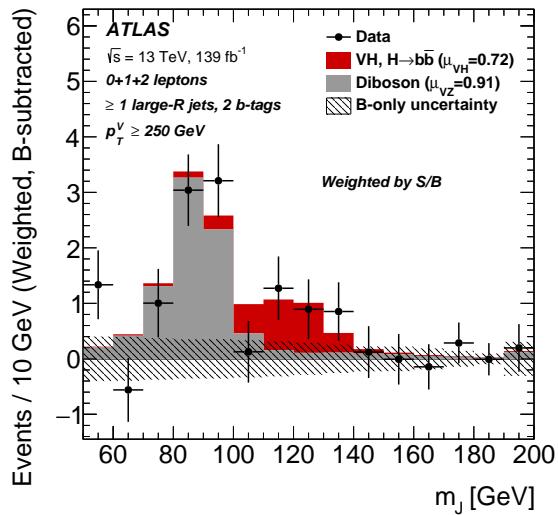
Alongside the standard 2-POI fit, a (3+1)-POI fit can be performed by splitting  $\mu_{VH}^{bb}$  into three separate POIs, one for each channel. A simultaneous fit to the channel



**Figure 7.6:** The  $m_J$  post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left)  $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$  and (right)  $p_T^V \geq 400 \text{ GeV}$ . The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions are shown as filled histograms. The Higgs boson signal ( $m_H = 125 \text{ GeV}$ ) is shown as a filled histogram and is normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and as an unstacked unfilled histogram, scaled by the SM prediction times a factor of two. The size of the combined on the sum of the fitted signal and background is shown in the hatched band. The highest bin contains the overflow [129].

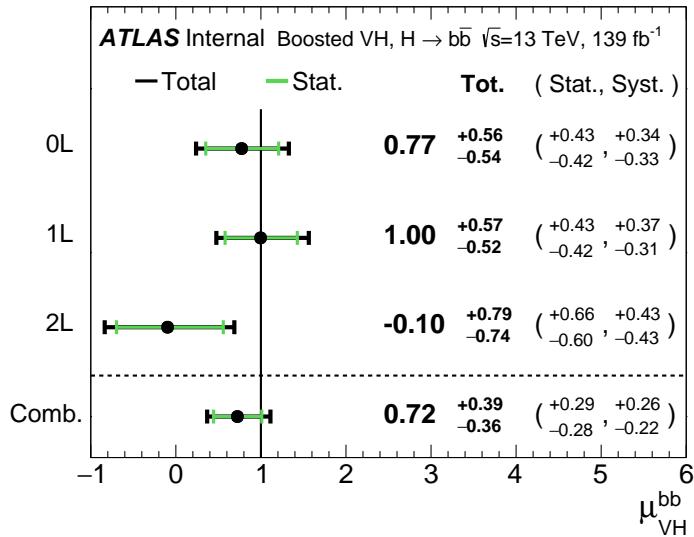


**Figure 7.7:** The  $m_J$  post-fit distributions in the  $t\bar{t}$  control region for (top) the 0-lepton channel and the 1-lepton channel for  $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$  and (bottom) the 0-lepton channel and the 1-lepton channel for  $p_T^V > 400 \text{ GeV}$ . The background contributions after the likelihood fit are shown as filled histograms. The Higgs boson signal ( $m_h = 125 \text{ GeV}$ ) is shown as a filled histogram on top of the fitted backgrounds normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and unstacked as an unfilled histogram, scaled by the SM prediction times a factor of 2. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The highest bin in the distributions contains the overflow [129].



**Figure 7.8:**  $m_J$  distribution in data after subtraction of all backgrounds except for the  $WZ$  and  $ZZ$  diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected contribution of the associated  $WH$  and  $ZH$  production of a SM Higgs boson with  $m_H = 125 \text{ GeV}$  is shown scaled by the measured combined signal strength ( $\mu_{VH}^{bb} = 0.72$ ). The diboson contribution is normalised to its best-fit value of  $\mu_{VZ}^{bb} = 0.91$ . The size of the combined statistical and systematic uncertainty is indicated by the hatched band. This error band is computed from a full signal-plus-background fit including all the systematic uncertainties defined in Section 7.2, except for the  $VH/VZ$  experimental and theory uncertainties [129].

specific signal strengths can then be performed, which allows a comparison of the contributions from each channel. Fig. 7.9 compares the best-fit signal strengths. The 0- and 1-lepton channels show a signal strength which is consistent with the SM prediction, while the 2-lepton channel shows a small deviation within the  $1\sigma$  uncertainty. Overall, good compatibility is observed via a  $\chi^2$  test with a corresponding  $p$ -value of 49%.



**Figure 7.9:** Signal strength compatibility test between the (3+1)-POI fit (with the three lepton channels splitted) and the default (1+1)-POI fit. The compatibility of the three channels is evaluated via a  $\chi^2$  difference test and results in a  $p$ -value of 49% [129].

#### 7.4.3 Impact of Systematics

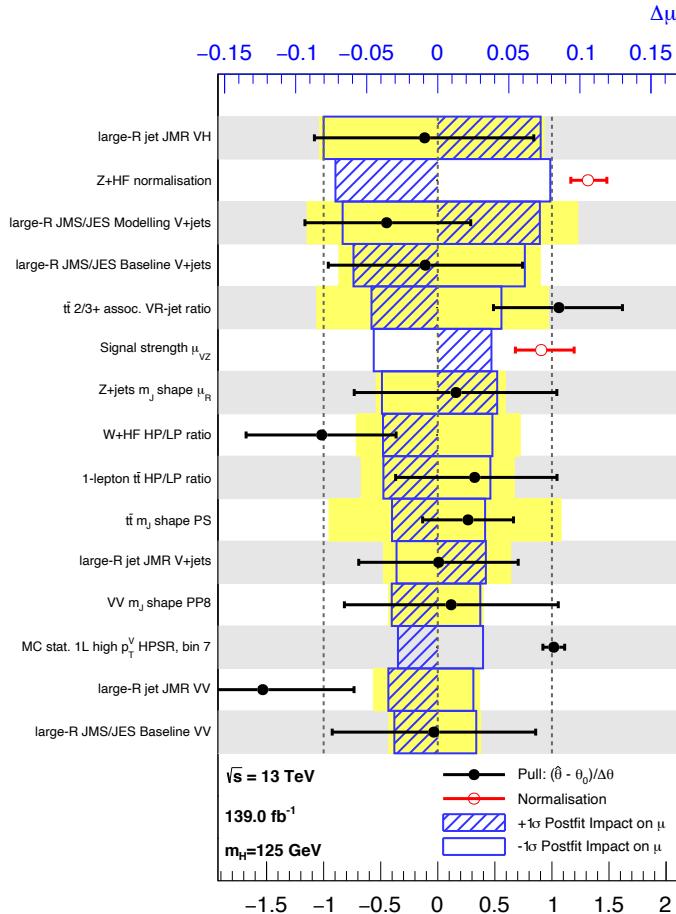
The impact of systematic uncertainties on the final fitted value  $\hat{\mu}^{bb}$  can be studied using the NP rankings, and the uncertainty breakdown.

Fig. 7.10 shows the NP ranking, which is used to visualise which out of the many NPs involved in the fit have the largest impact on the sensitivity to the fitted POI. To obtain the ranking, a likelihood scan is performed for each NP  $\theta_j$ . First, an unconditional fit is used to determine  $\hat{\theta}_j$ . The NP is then fixed to its post-fit value varied by  $\pm 1\sigma$ , the fit is repeated and the best-fit value of the POI,  $\Delta\hat{\mu}_{VH}^{bb}$ , is calculated, and used to rank the NPs. Also shown in Fig. 7.10 are the pulls and constraints for the highest ranked NPs.

2538 The experimental uncertainty on the signal large- $R$  jet mass resolution (JMR) has  
 2539 the largest impact. JMR and jet energy scale (JES) uncertainties also have impacts  
 2540 for the  $V$ +jets background and for the diboson background. The freely-floating  
 2541  $Z$ +hf normalisation is the second highest ranked NP, and is heavily constrained by  
 2542 the fit. The  $VZ$  POI  $\mu_{VZ}^{bb}$  is also a significant NP when considering the primary  $\mu_{VH}^{bb}$   
 2543 measurement.

2544 The NP ranking highlights individual NPs which have a large impact on the POI  
 2545 measurement sensitivity. Complementary information is provided at a higher level  
 2546 by considering the overall impact of different groups of systematics. The groups  
 2547 are constructed from NPs which have similar physical origin. The impact of each  
 2548 group is calculated by running a fit with all the NPs in the given group fixed to their  
 2549 nominal values. The uncertainty on the POI extracted from this fit is subtracted  
 2550 in quadrature from the the uncertainty on the POI from the nominal fit, and the  
 2551 resulting values are provided as the impact for each group. The full breakdown  
 2552 for the observed impact of uncertainties on the  $\mu_{VH}^{bb}$  signal strength is provided in  
 2553 Table 7.13. The total systematic impact is the difference in quadrature between  
 2554 the nominal uncertainty on  $\mu_{VH}^{bb}$  and the combined statistical impact. The “data  
 2555 stat only” group fixes all NPs at their nominal value, while the total statistical  
 2556 impact fixes all NPs except floating normalisations. The floating normalisations  
 2557 group fixes only the NPs associated with normalisation which are left floating in  
 2558 the fit. The uncertainty on  $\mu_{VH}^{bb}$  is dominated by combined statistcal effects (0.28),  
 2559 although the combined impact of systematics (0.24) is of a comparable size. The  
 2560 signal largest group is the data stat uncertainty (0.25), demonstrating that the  
 2561 analysis would benefit from an increased integrated luminosity or improved efficiency  
 2562 to select signal events (recall from Section 7.1.3 the signal efficiency is in the range of  
 2563 10%). Of the experimental systematic sources of uncertainty, the dominant impact  
 2564 is the experimental uncertainties associated with the reconstruction of large- $R$  jets  
 2565 (0.13). Other experimental sources of uncertainty are small in comparison. Modelling  
 2566 uncertainties also have a large contribution to the overall systematic uncertainty. The  
 2567 biggest contribution to the overall uncertainty is the combined statistical uncertainty  
 2568 on the simulated samples (0.09). Out of the backgrounds, the  $W$ +jets and  $Z$ +jets  
 2569 have the highest (0.06) and second-highest (0.05) impact respectively.  
 2570

Make sure it is clear in the text what all these correspond to, you might need to add a few more paragraphs on the main experimental uncertainties (which you should define).



**Figure 7.10:** Impact of systematic uncertainties on the fitted  $VH$  signal-strength parameter  $\hat{\mu}_{VH}^{\text{bb}}$  sorted in decreasing order. The boxes show the variations of  $\hat{\mu}$ , referring to the top  $x$ -axis, when fixing the corresponding individual nuisance parameter to its post-fit value modified upwards or downwards by its post-fit uncertainty, i.e.  $\hat{\theta} \pm \sigma_{\hat{\theta}}$ , and repeating the fit. The impact of up- and down-variations can be distinguished via the dashed and plane box fillings. The yellow boxes show the pre-fit impact (top  $x$ -axis) by varying each nuisance parameter by  $\pm 1$ . The filled circles show the deviation of the fitted value for each nuisance parameter,  $\hat{\theta}$ , from their nominal input value  $\theta_0$  expressed in standard deviations with respect to their nominal uncertainties  $\Delta\theta$  (bottom  $x$ -axis). The error bars show the post-fit uncertainties on  $\hat{\theta}$  with respect to their nominal uncertainties. The open circles show the fitted values and uncertainties of the normalization parameters that are freely floating in the fit. Pre-fit, these parameters have a value of one [174].

| Source of uncertainty                   | Signed impact      | Avg. impact     |
|---|--------------------|-----------------|
| Total                                   | +0.388 / -0.356    | 0.372           |
| Statistical                             | +0.286 / -0.280    | 0.283           |
| ↔ Data stat only                        | +0.251 / -0.245    | 0.248           |
| ↔ Floating normalisations               | +0.096 / -0.092    | 0.094           |
| Systematic                              | +0.261 / -0.219    | 0.240           |
| <hr/>                                   |                    |                 |
| Experimental uncertainties              |                    |                 |
| small-R jets                            | +0.041 / -0.034    | 0.038           |
| large-R jets                            | +0.161 / -0.105    | 0.133           |
| $E_T^{\text{miss}}$                     | +0.008 / -0.007    | 0.007           |
| Leptons                                 | +0.013 / -0.007    | 0.010           |
| <i>b</i> -tagging                       | <i>b</i> -jets     | +0.028 / -0.004 |
|   | <i>c</i> -jets     | +0.012 / -0.011 |
|   | light-flavour jets | +0.009 / -0.007 |
|   | extrapolation      | +0.004 / -0.005 |
| Pile-up                                 | +0.001 / -0.002    | 0.001           |
| Luminosity                              | +0.019 / -0.007    | 0.013           |
| <hr/>                                   |                    |                 |
| Theoretical and modelling uncertainties |                    |                 |
| Signal                                  | +0.073 / -0.026    | 0.050           |
| Backgrounds                             | +0.106 / -0.095    | 0.100           |
| ↔ $Z + \text{jets}$                     | +0.049 / -0.047    | 0.048           |
| ↔ $W + \text{jets}$                     | +0.059 / -0.056    | 0.058           |
| ↔ $t\bar{t}$                            | +0.037 / -0.032    | 0.035           |
| ↔ Single top quark                      | +0.031 / -0.023    | 0.027           |
| ↔ Diboson                               | +0.034 / -0.029    | 0.032           |
| ↔ Multijet                              | +0.009 / -0.009    | 0.009           |
| ↔ MC statistical                        | +0.091 / -0.092    | 0.092           |

**Table 7.13:** Breakdown of the absolute contributions to the uncertainty on the signal strength  $\mu_{VH}^{bb}$  obtained from the (1+1)-POI fit. The average impact represents the average between the positive and negative uncertainties on  $\mu_{VH}^{bb}$ . The sum in quadrature of the systematic uncertainties attached to the categories differs from the total systematic uncertainty due to correlations [174].

## 2574 7.5 Conclusion

2575 The analysis of the associated production of vector bosons with boosted Higgs bosons  
2576 decaying to a pair of  $b$ -quarks using large- $R$  jets is presented. The Higgs candidate is  
2577 reconstructed as a large- $R$  jet in order to improve sensitivity in the boosted regime  
2578 in which the Higgs decay products are significantly collimated. The analysis is  
2579 performed using  $139\text{ fb}^{-1}$  of proton–proton collision data at  $\sqrt{s} = 13\text{ TeV}$  collected  
2580 throughout the duration of Run 2 of the LHC.

2581 In comparison with the null hypothesis, the Standard Model (SM)  $VH$ ,  $H \rightarrow b\bar{b}$   
2582 process is found to have an observed significance of 2.1 standard deviations, whereas  
2583 the corresponding expected significance is 2.7 standard deviations. The  $VH$ ,  $H \rightarrow b\bar{b}$   
2584 process is measured simultaneously with the diboson  $VZ$ ,  $Z \rightarrow b\bar{b}$  process, which  
2585 provide a cross-check for the main analysis. The observed (expected) significance for  
2586 the diboson process is 5.4 (5.7).

2587 The statistical and systematic sources of uncertainty contribute a similar amount  
2588 to the overall uncertainty on the result. This analysis would therefore likely benefit  
2589 greatly from the improved  $b$ -tagging efficiency at high- $p_T$  enabled by GN1 as discussed  
2590 in Chapter 6, due to the associated reduction in statistical uncertainty provided by  
2591 the increased number of events used in the analysis.

Also mention something about the impact of the dominant systematics and how they could be reduced

# 2593 Chapter 8

## 2594 Conclusion

### 2595 8.1 Summary

2596 The current understanding of particle physics contains many unanswered questions,  
2597 and improving our understanding of the Standard Model is a promising way to  
2598 attempt to answer some of them. Key to this understanding is the Higgs Boson,  
2599 which was first observed only a decade ago and remains under intense scrutiny at the  
2600 LHC. Given it's propensity to decay to heavy flavour  $b$ -quarks, reconstructing and  
2601 identifying  $b$ -jets is of crucial importance to improving our understanding in this area.  
2602 As discussed in Chapter 4, this task becomes increasingly difficult at high transverse  
2603 momenta. The work in this thesis demonstrates that even with suboptimal track  
2604 reconstruction in this regime, it is possible to make algorithmic advancements to  
2605 the flavour tagging pipeline to improve the identification of  $b$ -jets. This work has  
2606 impacts for any analysis which relies on the identification of  $b$ -jets, including those  
2607 which are sensitive to the Higgs Boson.

2608 Analysis of  $VH$ ,  $H \rightarrow b\bar{b}$  events was also carried out with  $139\text{ fb}^{-1}$  of Run 2  
2609 ATLAS at  $\sqrt{s} = 13\text{ TeV}$ . Various background modelling uncertainties were derived  
2610 and investigations into the fit model were carried out. The analysis observed a  
2611 signal strength of  $\mu_{VH}^{bb} = 0.72_{-0.36}^{+0.39} = 0.72_{-0.28}^{+0.29}(\text{stat.})_{-0.22}^{+0.26}(\text{syst.})$  corresponding to an  
2612 observed (expected) significance of  $2.1\sigma$  ( $2.7\sigma$ ). The result was validated using a  
2613 simultaneous fit to the  $VZ$ ,  $Z \rightarrow b\bar{b}$  process.

## 2614 8.2 Future Work

2615 Further algorithmic improvements are likely to yield further improved flavour tagging  
2616 performance. Aside from these, large improvements to the flavour tagging perfor-  
2617 mance will likely be possible if improvements are made to the  $b$ -hadron decay track  
2618 reconstruction efficiency and accuracy.

2619 At the moment only the tracks from the Inner Detector and kinematic information  
2620 about the jet are provided as inputs to the tagging algorithms. In Chapter 6 it  
2621 was shown that the addition of a simple track-level variable corresponding to the  
2622 ID of the reconstruction lepton to the model improved the performance. However  
2623 there is still untapped potential in the form of additional information from the full  
2624 parameters of the reconstructed leptons (making full use of the Calorimeters and  
2625 Muon Spectrometer), the calorimeter clusters, and even the low level hits. Providing  
2626 such additional inputs to the model is likely to complement the information provided  
2627 by the tracks and further aid in the improvement of performane.

2628 On the output side, additional auxiliary training objectives may yield improved  
2629 performance and also help to add to the explainability of the model. Regression of  
2630 jet-level quantities such as the transverse momentum and mass, in addition to the  
2631 truth  $b$ -hadron decay length are promising regression targets.

2632 The GN1 architecture can also be readily optimised for new use cases and topologies,  
2633 as demonstreated by the studies described in Section 6.5. For example, a model  
2634 with only hit-level information provided as inputs could be used for a fast trigger  
2635 preselection on jets without the need to run the computationally expensive tracking  
2636 algorithms. The model could also be repurposed for primary vertexing, or a pile-up  
2637 jet tagger. Finally, the tagging of large- $R$  jets would benefit those analysis that rely  
2638 on it.

2639 Ultimately analysis which rely on the identification of heavy flavour jets will likely  
2640 benefit immensely from the improved performance of the flavour tagging algorithms.

2641 For example, the  $HH \rightarrow bbbb \dots$

make some  
claim about  
improved  
selection  
efficiency?

<sub>2642</sub>

# Bibliography

- <sub>2643</sub> [1] A. Buckley, *A class for typesetting academic theses*, (2010). <http://ctan.tug.org/tex-archive/macros/latex/contrib/heptesis/heptesis.pdf>.
- <sub>2644</sub>
- <sub>2645</sub> [2] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-027 (2022).
- <sub>2646</sub>
- <sub>2647</sub> [3] L. Morel, Z. Yao, P. Cladé and S. Guellati-Khélifa, *Determination of the fine-structure constant with an accuracy of 81 parts per trillion*, *Nature* **588** (2020) pp. 61–65.
- <sub>2648</sub>
- <sub>2649</sub>
- <sub>2650</sub> [4] T. Sailer, V. Debierre, Z. Harman, F. Heiße, C. König, J. Morgner et al., *Measurement of the bound-electron g-factor difference in coupled ions*, *Nature* **606** (2022) pp. 479–483.
- <sub>2651</sub>
- <sub>2652</sub>
- <sub>2653</sub> [5] CDF collaboration, *Observation of top quark production in  $\bar{p}p$  collisions*, *Phys. Rev. Lett.* **74** (1995) pp. 2626–2631 [[hep-ex/9503002](#)].
- <sub>2654</sub>
- <sub>2655</sub> [6] D0 collaboration, *Observation of the top quark*, *Phys. Rev. Lett.* **74** (1995) pp. 2632–2637 [[hep-ex/9503003](#)].
- <sub>2656</sub>
- <sub>2657</sub> [7] S. W. Herb et al., *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977) pp. 252–255.
- <sub>2658</sub>
- <sub>2659</sub> [8] UA1 collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $\sqrt{s} = 540$  GeV*, *Phys. Lett. B* **122** (1983) pp. 103–116.
- <sub>2660</sub>
- <sub>2661</sub>
- <sub>2662</sub> [9] DONUT collaboration, *Observation of tau neutrino interactions*, *Phys. Lett. B* **504** (2001) pp. 218–224 [[hep-ex/0012035](#)].
- <sub>2663</sub>
- <sub>2664</sub> [10] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) pp. 321–323.
- <sub>2665</sub>
- <sub>2666</sub> [11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.*

- 2667       **13** (1964) pp. 508–509.
- 2668       [12] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and*  
2669       *Massless Particles*, *Phys. Rev. Lett.* **13** (1964) pp. 585–587.
- 2670       [13] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*  
2671       *Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012)  
2672       p. 1 [[1207.7214](#)].
- 2673       [14] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*  
2674       *experiment at the LHC*, *Phys. Lett. B* **716** (2012) p. 30 [[1207.7235](#)].
- 2675       [15] Particle Data Group collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022)  
2676       p. 083C01.
- 2677       [16] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge*  
2678       *invariance*, *Phys. Rev.* **96** (1954) pp. 191–195.
- 2679       [17] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961)  
2680       pp. 579–588.
- 2681       [18] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) pp. 1264–1266.
- 2682       [19] A. Salam, *Weak and Electromagnetic Interactions*, *Proceedings of the 8th Nobel*  
2683       *symposium*, Ed. N. Svartholm, Almqvist & Wiksell, 1968, **Conf. Proc. C680519**  
2684       (1968) pp. 367–377.
- 2685       [20] T. D. Lee and C. N. Yang, *Question of parity conservation in weak interactions*,  
2686       *Phys. Rev.* **104** (1956) pp. 254–258.
- 2687       [21] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental*  
2688       *test of parity conservation in beta decay*, *Phys. Rev.* **105** (1957) pp. 1413–1415.
- 2689       [22] R. L. Garwin, L. M. Lederman and M. Weinrich, *Observations of the failure of*  
2690       *conservation of parity and charge conjugation in meson decays: the magnetic moment*  
2691       *of the free muon*, *Phys. Rev.* **105** (1957) pp. 1415–1417.
- 2692       [23] Y. Nambu, *Quasi-particles and gauge invariance in the theory of superconductivity*,  
2693       *Phys. Rev.* **117** (1960) pp. 648–663.
- 2694       [24] J. Goldstone, *Field theories with «superconductor» solutions*, *Il Nuovo Cimento*  
2695       (1955–1965) **19** (1961) pp. 154–164.
- 2696       [25] LHC Higgs Cross Section Working Group collaboration, *Handbook of LHC Higgs*

- 2697        *Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, [1610.07922](#).
- 2698 [26] ATLAS Collaboration, *Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector*, [ATLAS-CONF-2018-036](#) (2018).
- 2700 [27] CMS Collaboration, *Observation of Higgs Boson Decay to Bottom Quarks*, *Phys. Rev. Lett.* **121** (2018) p. 121801 [[1808.08242](#)].
- 2702 [28] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) p. S08001.
- 2703 [29] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) pp. S08002–S08002.
- 2705 [30] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) p. S08004.
- 2707 [31] The LHCb Collaboration, *The LHCb detector at the LHC*, *Journal of Instrumentation* **3** (2008) pp. S08005–S08005.
- 2709 [32] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) p. S08003.
- 2711 [33] ATLAS Collaboration, *Integrated luminosity summary plots for 2011-2012 data taking*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- 2714 [34] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, (2022). <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 2717 [35] CERN, “The accelerator complex.” (2012).
- 2718 [36] 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](#).
- 2721 [37] ATLAS Collaboration, *Atlas track reconstruction – general overview*, (2022). <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idooverview/>.
- 2723 [38] ATLAS Collaboration, *Performance of  $b$ -jet identification in the ATLAS experiment*, *JINST* **11** (2016) p. P04008 [[1512.01094](#)].
- 2725 [39] ATLAS Collaboration, *Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*, [ATLAS-CONF-2019-021](#) (2019).

- 2727 [40] ATLAS Collaboration, *ATLAS Detector and Physics Performance: Technical Design*  
2728 *Report, Volume 1*, .
- 2729 [41] for he ATLAS Collaboration collaboration, *Status of the ATLAS Experiment*, Tech.  
2730 Rep., Geneva (2010). 10.3204/DESY-PROC-2010-04/1.
- 2731 [42] ATLAS Collaboration, *ATLAS Inner Tracker Pixel Detector: Technical Design*  
2732 *Report*, ATLAS-TDR-030; CERN-LHCC-2017-021 (2017).
- 2733 [43] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design*  
2734 *Report*, ATLAS-TDR-025; CERN-LHCC-2017-005 (2017).
- 2735 [44] ATLAS Collaboration, *The inner detector*, (2022).  
2736 <https://atlas.cern/Discover/Detector/Inner-Detector>.
- 2737 [45] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*,  
2738 ATLAS-TDR-19; CERN-LHCC-2010-013 (2010).
- 2739 [46] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*,  
2740 *JINST* **13** (2018) p. T05008 [[1803.00844](https://arxiv.org/abs/1803.00844)].
- 2741 [47] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor*  
2742 *tracker*, *JINST* **9** (2014) p. P08009 [[1404.7473](https://arxiv.org/abs/1404.7473)].
- 2743 [48] ATLAS Collaboration, *Calorimeter*, (2022).  
2744 <https://atlas.cern/Discover/Detector/Calorimeter>.
- 2745 [49] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, .
- 2746 [50] G. Artoni, *Muon momentum scale and resolution in pp collisions at  $\sqrt{s} = 8 \text{ TeV}$  in*  
2747 *ATLAS*, Tech. Rep., Geneva (2016). 10.1016/j.nuclphysbps.2015.09.453.
- 2748 [51] ATLAS Collaboration, *Muon spectrometer*, (2022).  
2749 <https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
- 2750 [52] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur.*  
2751 *Phys. J. C* **77** (2017) p. 317 [[1611.09661](https://arxiv.org/abs/1611.09661)].
- 2752 [53] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*,  
2753 *ATL-SOFT-PUB-2021-001* (2021).
- 2754 [54] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig and E. Moyse, *Concepts,*  
2755 *Design and Implementation of the ATLAS New Tracking (NEWT)*, .
- 2756 [55] ATLAS Collaboration, *The Optimization of ATLAS Track Reconstruction in Dense*

- 2757       *Environments*, ATL-PHYS-PUB-2015-006 (2015).
- 2758 [56] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in*  
2759 *dense environments in LHC Run 2*, *Eur. Phys. J. C* **77** (2017) p. 673 [[1704.07983](#)].
- 2760 [57] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon*  
2761 *pixel detector*, *JINST* **9** (2014) p. P09009 [[1406.7690](#)].
- 2762 [58] R. Jansky, *Truth Seeded Reconstruction for Fast Simulation in the ATLAS*  
2763 *Experiment*, .
- 2764 [59] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment*  
2765 *in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) p. 332  
2766 [[1611.10235](#)].
- 2767 [60] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for*  
2768 *LHC Run 3*, ATL-PHYS-PUB-2019-015 (2019).
- 2769 [61] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency*  
2770 *measurement with  $t\bar{t}$  events in pp collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79**  
2771 (2019) p. 970 [[1907.05120](#)].
- 2772 [62] M. Cacciari, G. P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP*  
2773 **04** (2008) p. 063 [[0802.1189](#)].
- 2774 [63] M. Cacciari, G. P. Salam and G. Soyez, *Fastjet user manual*, *Eur.Phys.J. C***72** (2012)  
2775 p. 1896 [[1111.6097](#)].
- 2776 [64] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04**  
2777 (2008) p. 005 [[0802.1188](#)].
- 2778 [65] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its*  
2779 *performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) p. 490 [[1603.02934](#)].
- 2780 [66] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*  
2781 *the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) p. 466 [[1703.10485](#)].
- 2782 [67] ATLAS Collaboration, *Jet energy scale measurements and their systematic*  
2783 *uncertainties in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*,  
2784 *Phys. Rev. D* **96** (2017) p. 072002 [[1703.09665](#)].
- 2785 [68] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet Substructure as a*  
2786 *New Higgs Search Channel at the Large Hadron Collider*, *Phys. Rev. Lett.* **100** (2008)  
2787 p. 242001 [[0802.2470](#)].

- 2788 [69] ATLAS collaboration, *Measurement of the associated production of a Higgs boson*  
2789 *decaying into b-quarks with a vector boson at high transverse momentum in pp*  
2790 *collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)  
2791 p. 136204 [[2008.02508](#)].
- 2792 [70] ATLAS collaboration, *Performance of jet substructure techniques for large-R jets in*  
2793 *proton-proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *JHEP* **09** (2013)  
2794 p. 076 [[1306.4945](#)].
- 2795 [71] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in*  
2796 *13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019)  
2797 p. 135 [[1807.09477](#)].
- 2798 [72] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, *JHEP* **06** (2009) p. 059  
2799 [[0903.0392](#)].
- 2800 [73] ATLAS Collaboration, *Variable Radius, Exclusive- $k_T$ , and Center-of-Mass Subjet*  
2801 *Reconstruction for Higgs( $\rightarrow b\bar{b}$ ) Tagging in ATLAS*, [ATL-PHYS-PUB-2017-010](#)  
2802 (2017).
- 2803 [74] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the*  
2804 *Gaussian Sum Filter-based model for bremsstrahlung*, [ATLAS-CONF-2012-047](#) (2012).
- 2805 [75] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector*  
2806 *using the 2015 LHC proton–proton collision data*, [ATLAS-CONF-2016-024](#) (2016).
- 2807 [76] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in*  
2808 *proton–proton collision data at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **76** (2016) p. 292  
2809 [[1603.05598](#)].
- 2810 [77] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*  
2811 *with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys.*  
2812 *J. C* **78** (2018) p. 903 [[1802.08168](#)].
- 2813 [78] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS*  
2814 *b-tagging algorithms for the 2017-18 LHC run*, [ATL-PHYS-PUB-2017-013](#) (2017).
- 2815 [79] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the*  
2816 *ATLAS detector*, [ATL-PHYS-PUB-2017-011](#) (2017).
- 2817 [80] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*  
2818 *Neural Networks at the ATLAS Experiment*, [ATL-PHYS-PUB-2017-003](#) (2017).

- 2819 [81] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter*  
2820 *Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014 (2020).
- 2821 [82] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in*  
2822 *Run-2*, ATL-PHYS-PUB-2015-022 (2015).
- 2823 [83] B. R. Webber, *Fragmentation and hadronization*, *Int. J. Mod. Phys. A* **15S1** (2000)  
2824 pp. 577–606 [[hep-ph/9912292](#)].
- 2825 [84] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98**  
2826 (2018) p. 030001.
- 2827 [85] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom*  
2828 *and charm hadrons in the decays of top quarks and the fragmentation of high  $p_T$  jets*,  
2829 ATL-PHYS-PUB-2014-008 (2014).
- 2830 [86] Nazar Bartosik, *Diagram showing the common principle of identification of jets*  
2831 *initiated by b-hadron decays*, (2022).  
2832 [https://en.m.wikipedia.org/wiki/File:B-tagging\\_diagram.png](https://en.m.wikipedia.org/wiki/File:B-tagging_diagram.png).
- 2833 [87] ATLAS Collaboration, *Tracking efficiency studies in dense environments*, (2022).  
2834 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2022-03/>.
- 2835 [88] S. Adorni Braccesi Chiassi, *Vector Boson Tagging in the Context of the Search for*  
2836 *New Physics in the Fully Hadronic Final State*, Ph.D. thesis, Geneva U., 2021.  
2837 10.13097/archive-ouverte/unige:158558.
- 2838 [89] E. Boos, M. Dobbs, W. Giele, I. Hinchliffe, J. Huston, V. Ilyin et al., *Generic User*  
2839 *Process Interface for Event Generators*, ArXiv High Energy Physics - Phenomenology  
2840 *e-prints* (2001) .
- 2841 [90] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley et al., *A*  
2842 *standard format for Les Houches Event Files*, *Computer Physics Communications*  
2843 **176** (2007) pp. 300–304.
- 2844 [91] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are*  
2845 *universal approximators*, *Neural Networks* **2** (1989) pp. 359–366.
- 2846 [92] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning representations by*  
2847 *back-propagating errors*, *nature* **323** (1986) pp. 533–536.
- 2848 [93] W. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous*  
2849 *activity*, *The bulletin of mathematical biophysics* **5** (1943) pp. 115–133.

- 2850 [94] J. Hopfield, *Neural networks and physical systems with emergent collective*  
2851 *computational abilities*, in *Spin Glass Theory and Beyond: An Introduction to the*  
2852 *Replica Method and Its Applications*, pp. 411–415, World Scientific, (1987).
- 2853 [95] Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, *Efficient backprop*, in *Neural*  
2854 *networks: Tricks of the trade*, pp. 9–48, Springer, (2012).
- 2855 [96] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv*  
2856 *e-prints* (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2857 [97] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi,  
2858 M. Malinowski et al., *Relational inductive biases, deep learning, and graph networks*,  
2859 *arXiv e-prints* (2018) p. arXiv:1806.01261 [[1806.01261](#)].
- 2860 [98] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R. Salakhutdinov and A. Smola,  
2861 *Deep Sets*, *arXiv e-prints* (2017) p. arXiv:1703.06114 [[1703.06114](#)].
- 2862 [99] P. Nason, *A new method for combining nlo qcd with shower monte carlo algorithms*,  
2863 *Journal of High Energy Physics* **2004** (2004) p. 040–040.
- 2864 [100] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order monte*  
2865 *carlo for heavy flavour hadroproduction*, *Journal of High Energy Physics* **2007** (2007)  
2866 p. 126–126.
- 2867 [101] S. Frixione, P. Nason and C. Oleari, *Matching nlo qcd computations with parton*  
2868 *shower simulations: the powheg method*, *Journal of High Energy Physics* **2007** (2007)  
2869 p. 070–070.
- 2870 [102] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing nlo*  
2871 *calculations in shower monte carlo programs: the powheg box*, *Journal of High Energy*  
2872 *Physics* **2010** (2010) .
- 2873 [103] NNPDF collaboration, *Parton distributions for the LHC run II*, *JHEP* **04** (2015)  
2874 p. 040 [[1410.8849](#)].
- 2875 [104] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,  
2876 *ATL-PHYS-PUB-2016-020* (2016).
- 2877 [105] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An*  
2878 *introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) p. 159  
2879 [[1410.3012](#)].
- 2880 [106] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*,

- 2881        ATL-PHYS-PUB-2014-021 (2014).
- 2882 [107] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013)  
2883        p. 244 [[1207.1303](#)].
- 2884 [108] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A*  
2885        **462** (2001) p. 152.
- 2886 [109] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**  
2887        (2010) p. 823 [[1005.4568](#)].
- 2888 [110] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4 – a simulation toolkit*, *Nucl.*  
2889        *Instrum. Meth. A* **506** (2003) p. 250.
- 2890 [111] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS*  
2891        *detector*, [ATLAS-CONF-2014-018](#) (2014).
- 2892 [112] ATLAS Collaboration, *Observation of  $H \rightarrow b\bar{b}$  decays and VH production with the*  
2893        *ATLAS detector*, *Phys. Lett. B* **786** (2018) p. 59 [[1808.08238](#)].
- 2894 [113] ATLAS Collaboration, *Observation of Higgs boson production in association with a*  
2895        *top quark pair at the LHC with the ATLAS detector*, *Phys. Lett. B* **784** (2018) p. 173  
2896        [[1806.00425](#)].
- 2897 [114] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs*  
2898        *using  $139\text{ fb}^{-1}$  of  $pp$  collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector*, *JHEP* **03**  
2899        (2020) p. 145 [[1910.08447](#)].
- 2900 [115] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky et al.,  
2901        *Secondary vertex finding in jets with neural networks*, *The European Physical Journal*  
2902        *C* **81** (2021) .
- 2903 [116] H. Serviansky, N. Segol, J. Shlomi, K. Cranmer, E. Gross, H. Maron et al.,  
2904        *Set2Graph: Learning Graphs From Sets*, *arXiv e-prints* (2020) p. arXiv:2002.08772  
2905        [[2002.08772](#)].
- 2906 [117] ATLAS Collaboration, *Muon reconstruction performance in early  $\sqrt{s} = 13\text{ TeV}$  data*,  
2907        [ATL-PHYS-PUB-2015-037](#) (2015).
- 2908 [118] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS*  
2909        *experiment using the 2015 and 2016 LHC proton–proton collision data at*  
2910         *$\sqrt{s} = 13\text{ TeV}$* , *Eur. Phys. J. C* **79** (2019) p. 639 [[1902.04655](#)].
- 2911 [119] D. Hwang, J. Park, S. Kwon, K.-M. Kim, J.-W. Ha and H. J. Kim, *Self-supervised*

- 2912       *Auxiliary Learning with Meta-paths for Heterogeneous Graphs*, arXiv e-prints (2020)  
2913       p. arXiv:2007.08294 [[2007.08294](#)].
- 2914     [120] S. Brody, U. Alon and E. Yahav, *How Attentive are Graph Attention Networks?*,  
2915       arXiv e-prints (2021) p. arXiv:2105.14491 [[2105.14491](#)].
- 2916     [121] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez et al.,  
2917       *Attention Is All You Need*, arXiv e-prints (2017) p. arXiv:1706.03762 [[1706.03762](#)].
- 2918     [122] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv  
2919       e-prints (2014) p. arXiv:1412.6980 [[1412.6980](#)].
- 2920     [123] J. Bai, F. Lu, K. Zhang et al., *ONNX: Open neural network exchange*, (2019).  
2921       <https://github.com/onnx/onnx>.
- 2922     [124] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp*  
2923       *collision dataset*, [2211.16345](#).
- 2924     [125] ATLAS Collaboration, *Performance of Run 3 HLT b-tagging with fast tracking,*  
2925       *Public b-Jet Trigger Plots for Collision Data*, (2022).  
2926       [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults#Performance_of_Run_3_HLT_b_taggi)  
2927       [BJetTriggerPublicResults#Performance\\_of\\_Run\\_3\\_HLT\\_b\\_taggi](#).
- 2928     [126] ATLAS Collaboration, *Neural Network Jet Flavour Tagging with the Upgraded*  
2929       *ATLAS Inner Tracker Detector at the High-Luminosity LHC*,  
2930       [ATL-PHYS-PUB-2022-047](#) (2022).
- 2931     [127] R. Lafaye, T. Plehn, M. Rauch, D. Zerwas and M. Duhrssen, *Measuring the Higgs*  
2932       *Sector*, *JHEP* **08** (2009) p. 009 [[0904.3866](#)].
- 2933     [128] ATLAS Collaboration, *Measurements of WH and ZH production in the  $H \rightarrow b\bar{b}$*   
2934       *decay channel in pp collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C*  
2935       **81** (2021) p. 178 [[2007.02873](#)].
- 2936     [129] ATLAS Collaboration, *Measurement of the associated production of a Higgs boson*  
2937       *decaying into b-quarks with a vector boson at high transverse momentum in pp*  
2938       *collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021)  
2939       p. 136204 [[2008.02508](#)].
- 2940     [130] K. Mimasu, V. Sanz and C. Williams, *Higher order QCD predictions for associated*  
2941       *Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) p. 039  
2942        [[1512.02572](#)].

- 2943 [131] ATLAS Collaboration, *Performance of jet substructure techniques for large- $R$  jets in*  
 2944 *proton–proton collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, *JHEP* **09** (2013)  
 2945 p. 076 [[1306.4945](#)].
- 2946 [132] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO*  
 2947 *calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010)  
 2948 p. 043 [[1002.2581](#)].
- 2949 [133] ATLAS collaboration, *Measurement of the  $Z/\gamma^*$  boson transverse momentum*  
 2950 *distribution in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **09**  
 2951 (2014) p. 145 [[1406.3660](#)].
- 2952 [134] G. Cullen, N. Greiner, G. Heinrich, G. Luisoni, P. Mastrolia, G. Ossola et al.,  
 2953 *Automated one-loop calculations with GoSam*, *Eur. Phys. J. C* **72** (2012) p. 1889  
 2954 [[1111.2034](#)].
- 2955 [135] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: multi-scale improved NLO*,  
 2956 *JHEP* **10** (2012) p. 155 [[1206.3572](#)].
- 2957 [136] G. Luisoni, P. Nason, C. Oleari and F. Tramontano,  *$HW^\pm/HZ + 0$  and 1 jet at*  
 2958 *NLO with the POWHEG BOX interfaced to GoSam and their merging within*  
 2959 *MiNLO*, *JHEP* **10** (2013) p. 083 [[1306.2542](#)].
- 2960 [137] M. L. Ciccolini, S. Dittmaier and M. Krämer, *Electroweak radiative corrections to*  
 2961 *associated  $WH$  and  $ZH$  production at hadron colliders*, *Phys. Rev. D* **68** (2003)  
 2962 p. 073003 [[hep-ph/0306234](#)].
- 2963 [138] O. Brein, A. Djouadi and R. Harlander, *NNLO QCD corrections to the*  
 2964 *Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) pp. 149–156  
 2965 [[hep-ph/0307206](#)].
- 2966 [139] G. Ferrera, M. Grazzini and F. Tramontano, *Associated Higgs-W-Boson Production*  
 2967 *at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys. Rev. Lett.*  
 2968 **107** (2011) p. 152003 [[1107.1164](#)].
- 2969 [140] O. Brein, R. V. Harlander, M. Wiesemann and T. Zirke, *Top-quark mediated effects*  
 2970 *in hadronic Higgs-Strahlung*, *Eur. Phys. J. C* **72** (2012) p. 1868 [[1111.0761](#)].
- 2971 [141] G. Ferrera, M. Grazzini and F. Tramontano, *Higher-order QCD effects for associated*  
 2972  *$WH$  production and decay at the LHC*, *JHEP* **04** (2014) p. 039 [[1312.1669](#)].
- 2973 [142] G. Ferrera, M. Grazzini and F. Tramontano, *Associated  $ZH$  production at hadron*  
 2974 *colliders: The fully differential NNLO QCD calculation*, *Phys. Lett. B* **740** (2015)

- 2975 pp. 51–55 [[1407.4747](#)].
- 2976 [143] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson*  
2977 *at NNLO*, *JHEP* **06** (2016) p. 179 [[1601.00658](#)].
- 2978 [144] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,  
2979 *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **02** (2013) p. 078  
2980 [[1211.5015](#)].
- 2981 [145] B. Hespel, F. Maltoni and E. Vryonidou, *Higgs and Z boson associated production via*  
2982 *gluon fusion in the SM and the 2HDM*, *JHEP* **06** (2015) p. 065 [[1503.01656](#)].
- 2983 [146] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke, *Soft gluon resummation for*  
2984 *gluon-induced Higgs Strahlung*, *JHEP* **11** (2014) p. 082 [[1410.0217](#)].
- 2985 [147] R. V. Harlander, S. Liebler and T. Zirke, *Higgs Strahlung at the Large Hadron*  
2986 *Collider in the 2-Higgs-doublet model*, *JHEP* **02** (2014) p. 023 [[1307.8122](#)].
- 2987 [148] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo – Higgs Strahlung at hadron*  
2988 *colliders*, *Comput. Phys. Commun.* **184** (2013) pp. 998–1003 [[1210.5347](#)].
- 2989 [149] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte*  
2990 *Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) p. 126 [[0707.3088](#)].
- 2991 [150] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair*  
2992 *cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) p. 2930  
2993 [[1112.5675](#)].
- 2994 [151] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with*  
2995 *shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) p. 111  
2996 [[0907.4076](#)].
- 2997 [152] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top*  
2998 *quark production*, *Phys. Rev. D* **81** (2010) p. 054028 [[1001.5034](#)].
- 2999 [153] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for*  
3000 *t-channel single top quark production*, *Phys. Rev. D* **83** (2011) p. 091503 [[1103.2792](#)].
- 3001 [154] E. Re, *Single-top Wt-channel production matched with parton showers using the*  
3002 *POWHEG method*, *Eur. Phys. J. C* **71** (2011) p. 1547 [[1009.2450](#)].
- 3003 [155] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated*  
3004 *production with a  $W^-$  or  $H^-$* , *Phys. Rev. D* **82** (2010) p. 054018 [[1005.4451](#)].

- 3005 [156] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al.,  
3006     *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) p. 007 [[0811.4622](#)].
- 3007 [157] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) p. 034  
3008     [[1905.09127](#)].
- 3009 [158] F. Caccioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,  
3010     *Phys. Rev. Lett.* **108** (2012) p. 111601 [[1111.5206](#)].
- 3011 [159] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008)  
3012     p. 039 [[0808.3674](#)].
- 3013 [160] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour  
3014 dipole factorisation*, *JHEP* **03** (2008) p. 038 [[0709.1027](#)].
- 3015 [161] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton  
3016 showers. The NLO case*, *JHEP* **04** (2013) p. 027 [[1207.5030](#)].
- 3017 [162] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector Boson  
3018 Production at Hadron Colliders: a Fully Exclusive QCD Calculation at NNLO*, *Phys.  
3019 Rev. Lett.* **103** (2009) p. 082001 [[0903.2120](#)].
- 3020 [163] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys. G* **43**  
3021     (2016) p. 023001 [[1510.03865](#)].
- 3022 [164] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *Electroweak corrections to  
3023 Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk*, *JHEP* **03**  
3024     (2012) p. 075 [[1112.5142](#)].
- 3025 [165] A. Denner, S. Dittmaier, S. Kallweit and A. Mück, *HAWK 2.0: A Monte Carlo  
3026 program for Higgs production in vector-boson fusion and Higgs strahlung at hadron  
3027 colliders*, *Comput. Phys. Commun.* **195** (2015) pp. 161–171 [[1412.5390](#)].
- 3028 [166] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS  
3029 detector using t̄t events at √s = 13 TeV*, *JHEP* **08** (2018) p. 089 [[1805.01845](#)].
- 3030 [167] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using  
3031 ATLAS proton–proton collision data at √s = 13 TeV*, *ATLAS-CONF-2018-006*  
3032     (2018).
- 3033 [168] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in t̄t events using  
3034 a likelihood approach with the ATLAS detector*, *ATLAS-CONF-2018-001* (2018).
- 3035 [169] ATLAS Collaboration, *Flavor Tagging with Track-Jets in Boosted Topologies with the*

- 3036        *ATLAS Detector*, ATL-PHYS-PUB-2014-013 (2014).
- 3037 [170] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark*  
3038        *pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) p. 836  
3039        [[1906.11005](#)].
- 3040 [171] J. K. Anders and M. D'Onofrio, *V+Jets theoretical uncertainties estimation via a*  
3041        *parameterisation method*, Tech. Rep. ATL-COM-PHYS-2016-044, Geneva (2016).
- 3042 [172] L. A. Harland-Lang, A. D. Martin, P. Motylinski and R. S. Thorne, *Parton*  
3043        *distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015) p. 204  
3044        [[1412.3989](#)].
- 3045 [173] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky et al., *New parton*  
3046        *distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev.*  
3047        **D93** (2016) p. 033006 [[1506.07443](#)].
- 3048 [174] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in*  
3049        *association with a vector boson and decaying to a pair of b-quarks using large-R jets*,  
3050        ATL-COM-PHYS-2019-1125 (2019).
- 3051 [175] J. Pumplin et al., *New generation of parton distributions with uncertainties from*  
3052        *global qcd analysis*, *JHEP* **07** (2002) p. 012 [[0201195](#)].
- 3053 [176] L. Moneta, K. Belasco, K. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo et al., *The*  
3054        *roostats project*, *arXiv preprint arXiv:1009.1003* (2010) .

# 3055 List of figures

|                     |   |                        |
|---------------------|---|------------------------|
| <small>3056</small> | 2.1 The Higgs potential $V(\phi)$ of the complex scalar field singlet $\phi = \phi_1 + i\phi_2$ , with a choice of $\mu^2 < 0$ leading to a continuous degeneracy in the true vacuum states. A false vacuum is present at the origin. The SM Higgs mechanism relies on a complex scalar doublet with a corresponding 5-dimensional potential. . . . . | <small>3057</small> 20 |
| <small>3061</small> | 2.2 Diagrams for the four main Higgs boson production modes at the LHC for a Higgs mass $m_H = 125$ GeV at a centre of mass energy $\sqrt{s} = 13$ TeV. . . . .   | <small>3062</small> 26 |
| <small>3063</small> | 2.3 Higgs boson production cross sections as a function of Higgs mass ( $m_H$ ) at $\sqrt{s} = 13$ TeV [25]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, Higgs boson production is dominated by gluon-gluon fusion, vector boson fusion, associated production with vector bosons, and top quark fusion. . . . .                 | <small>3064</small> 26 |
| <small>3068</small> | 2.4 Higgs boson branching ratios as a function of Higgs mass ( $m_H$ ) at $\sqrt{s} = 13$ TeV [25]. Uncertainties are shown in the shaded bands. At $m_H = 125$ GeV, the Higgs predominantly decays to a pair of $b$ -quarks, around 58% of the time. The leading subdominant decay mode is to a pair of $W$ bosons. . . . .                          | <small>3069</small> 27 |
| <small>3072</small> | 3.1 An overview of the CERN accelerator complex [35]. The LHC is fed by a series of accelerators starting with Linac4. Next are the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron which injects protons into the LHC. . . . .  | <small>3073</small> 30 |
| <small>3076</small> | 3.2 The coordinate system used at the ATLAS detector, showing the locations of the four main experiments located at various points around the LHC. The 3-vector momentum $\mathbf{p} = (p_x, p_y, p_z)$ is shown by the red arrow. Reproduced from Ref. [36]. . . . .   | <small>3077</small> 31 |
| <small>3078</small> |   | <small>3079</small>    |

|      |  |    |
|------|--|----|
| 3080 | 3.3 The track parameterisation used at the ATLAS detector. Five coordinates ( $d_0, z_0, \phi, \theta, q/p$ ) are specified, defined at the track's point of closest approach to the nominal interaction point at the origin of the coordinate system. The figure shows the three-momentum $\mathbf{p}$ and the transverse momentum $p_T$ (defined in Eq. (3.2)). The basis vectors $e_x$ , $e_y$ and $e_z$ are also shown. Reproduced from Ref. [37]. . . . . | 33 |
| 3086 | 3.4 Delivered, recorded, and usable integrated luminosity as a function of time over the course of Run 2 [34]. A total of $139\text{ fb}^{-1}$ of collision data is labelled as good for physics, meaning all sub-detector systems were operating nominally. . . . .   | 35 |
| 3090 | 3.5 Average pile-up profiles measured by ATLAS during Run 2 [34]. Higher levels of pile-up are planned for Run 3. . . . .  | 35 |
| 3092 | 3.6 A 3D model of the entire ATLAS detector [41]. The detector is 46 m long and 25 m in diameter. Cutouts to the centre of the detector reveal the different subdetectors which are arranged in concentric layers around the nominal interaction point. . . . .  | 36 |
| 3096 | 3.7 A 3D model of the ATLAS ID, made up of the pixel and SCT subdetectors, showing the barrel layers and end-cap disks [44]. . . . .   | 37 |
| 3098 | 3.8 A cross-sectional view of the ATLAS ID, with the radii of the different barrel layers shown [37]. . . . .  | 38 |
| 3100 | 3.9 A schematic cross-sectional view of the ATLAS IBL [45]. . . . .  | 39 |
| 3101 | 3.10 The ATLAS calorimeters [48]. The ECal uses LAr-based detectors, while the HCal uses mainly scintillating tile detectors. In the forward region the HCal also includes the LAr hadronic endcaps. . . . .   | 41 |
| 3104 | 3.11 The ATLAS muon spectrometer [51]. . . . .   | 43 |
| 3105 | 3.12 Particles (left) which have enough separation will leave charge depositions which are resolved into separate clusters. Sufficiently close particles (right) can lead to merged clusters. Their combined energy deposits are reconstructed as a single merged cluster [56]. . . . .  | 47 |
| 3109 | 3.13 A sketch of electron reconstruction using the ATLAS detector [75]. Electron reconstruction makes use of the entire ID and the calorimeters. In particular, discriminating information comes from the TRT and ECal. . . . .  | 51 |

|      |   |    |
|------|---|----|
| 3112 | 4.1 The truth $b$ -hadron decay radius $L_{xy}$ as a function of truth transverse<br>3113 momentum $p_T$ for reconstructed $b$ -jets in $Z'$ events. Error bars show the<br>3114 standard deviation. The pixel layers are shown in dashed horizontal lines.   | 55 |
| 3115 | 4.2 Diagram of a typical $b$ -jet (blue) which has been produced in an event<br>3116 alongside two light jets (grey) [86]. The $b$ -hadron has travelled a significant<br>3117 distance (pink dashed line) from the primary interaction point (pink dot)<br>3118 before its decay. The large transverse impact parameter $d_0$ is a characteristic<br>3119 property of the trajectories of $b$ -hadron decay products. . . . .  | 56 |
| 3120 | 4.3 At lower $p_T$ (left) the decay length of the $b$ -hadron is reduced, and the<br>3121 resulting decay tracks are less collimated. At higher $p_T$ (right) the $b$ -hadron<br>3122 decay length increases and the resulting decay tracks are more collimated and<br>3123 have less distance over which to diverge before reaching detector elements.<br>3124 As a result, the ID may be unable to resolve charged depositions from<br>3125 different particles, resulting merged clusters. . . . .   | 57 |
| 3126 | 4.4 $b$ -hadron decay track reconstruction efficiency as a function of truth $b$ -hadron<br>3127 $p_T$ [87]. Nominal track reconstruction is shown in black, while the track<br>3128 reconstruction efficiency for track candidates (i.e. the pre-ambiguity solver<br>3129 efficiency) is shown in green. For high- $p_T$ $b$ -hadrons, the ambiguity solver<br>3130 is overly aggressive in its removal of $b$ -hadron decay tracks. Suggestions for<br>3131 the improvement of the track reconstruction efficiency in this regime by the<br>3132 loosening of cuts in the ambiguity solver are shown in blue and red. . . . .   | 58 |
| 3133 | 4.5 Hit multiplicities on the IBL (left) and the pixel layers (right) as a function<br>3134 of the $p_T$ of the reconstructed track. Tracks from the weak decay of<br>3135 the $b$ -hadron are shown in red, while fragmentation tracks (which are<br>3136 prompt) are in blue. Baseline tracks are those produced in the standard<br>3137 reconstruction described in Section 3.4.1, while pseudotrack represent the<br>3138 ideal performance of the ATLAS detector and are described in Section 3.4.1.<br>3139 Hit multiplicities on the pseudotracks decrease at high $p_T$ due to the flight<br>3140 of the $b$ -hadron before its decay. The baseline tracks have more hits than the<br>3141 pseudotracks, indicating that they are being incorrectly assigned additional<br>3142 hits on the inner layers of the detector. . . . . | 59 |
| 3143 | 4.6 The fraction of hits which are shared on $b$ -hadron decay tracks on the IBL<br>3144 (left), and the B-layer (right), as a function of the production radius of<br>3145 the $b$ -hadron decay product. Pseudotracks represent the ideal performance<br>3146 given the ATLAS detector, see Section 3.4.1. . . . .  | 61 |

---

|      |  |    |
|------|--|----|
| 3147 | 4.7 The rate to assign good (left) and wrong (right) IBL hits as a function<br>3148 of $b$ -hadron $p_T$ for tracks using baseline tracking (black) and the modified<br>3149 version of the outlier removal procedure (red). For each track, the corre-<br>3150 sponding $p_T$ bin is filled with the number of good or wrong hits and this<br>3151 value is averaged to show the overall rate. . . . .  | 63 |
| 3152 | 4.8 (left) $b$ -hadron decay track $d_0$ pulls ( $d_0/s(d_0)$ ) for baseline and modified<br>3153 GX2F tracks. (right) The absolute value of the $d_0$ pull as a function of<br>3154 $b$ -hadron transverse momentum. . . . .  | 64 |
| 3155 | 5.1 A diagram displaying the logical flow of a single neuron with three inputs<br>3156 $x_i^j$ . Each input is multiplied by a weight $\theta_j$ , and the resulting values are<br>3157 summed. A bias term $\theta_0$ is added, and the result $z$ is passed to an activation<br>3158 function. Each neuron can be thought of as a logistic regression model. .   | 67 |
| 3159 | 5.2 The output of several common choices for the activation function $g(z)$ of<br>3160 an artificial neuron. The input $z$ is the output of the dot product between<br>3161 the activation and the weights, plus a bias term. . . . .  | 68 |
| 3162 | 5.3 The flow of information through a graph neural network layer for (left) a full<br>3163 implementation of the layer and (right) a deep sets model [98]. Reproduced<br>3164 from Ref. [97]. . . . .  | 71 |
| 3165 | 5.4 Rate of fake tracks as a function of jet transverse momentum (left) and<br>3166 $\Delta R(\text{track}, \text{jet})$ (right). The rate of fake tracks increases significantly as a<br>3167 function of $p_T$ , and also increases as the distance to the jet axis decreases.   | 73 |
| 3168 | 5.5 The light-jet efficiency of the low level tagger SV1 for jets in the $Z'$ sample<br>3169 with $250 < p_T < 5000$ GeV, as a function of $b$ -jet efficiency. The nominal<br>3170 tracking setup (black) is shown alongside the case where fake tracks which<br>3171 are not from the decay of a $b$ -hadron are removed. The light-jet efficiency<br>3172 is decreased, demonstrating that the presence of fake tracks is detrimental<br>3173 to the algorithm performance. . . . . | 73 |

---

|      |   |    |
|------|---|----|
| 3174 | 5.6 (left) Normalised histogram of the model output separated for signal and<br>3175 fake tracks, and further separated by those tracks which are from the<br>3176 decay of a $b$ -hadron. (right) The ROC curve for all tracks (solid line) and<br>3177 tracks from the decay of a $b$ -hadron (dashed line). The model is able to<br>3178 successfully discriminate between signal and fake tracks, and shows a very<br>3179 similar performance when looking specifically at tracks from the decay of a<br>3180 $b$ -hadron. . . . .   | 77 |
| 3181 | 5.7 (left) Normalised histogram of the model output separated for tracks from<br>3182 the decay of a $b$ -hadron and tracks from other sources. The two groups are<br>3183 further separated into those tracks which are fake. (right) The ROC curve<br>3184 for all tracks (solid line). . . . .   | 79 |
| 3185 | 5.8 The effect of applying the fake track identification algorithm together with<br>3186 the $b$ -hadron decay track identification on the jet tagging performance of<br>3187 SV1 for jets in the $Z'$ sample with $250 \text{ GeV} < p_T < 400 \text{ GeV}$ (left) and<br>3188 $400 \text{ GeV} < p_T < 1 \text{ TeV}$ (right). The nominal SV1 light-jet rejection (black)<br>3189 is compared to two working points of fake track removal, labelled “A” (red)<br>3190 and “B” (orange), which represent two different 2D working points of the<br>3191 track classification tools. Removal of fake tracks based on truth information<br>3192 is shown by the green curves. . . . . | 80 |
| 3193 | 6.1 Comparison of the existing flavour tagging scheme (left) and GN1 (right) [2].<br>3194 The existing approach utilises low-level algorithms (shown in blue), the<br>3195 outputs of which are fed into a high-level algorithm (DL1r). Instead<br>3196 of being used to guide the design of the manually optimised algorithms,<br>3197 additional truth information from the simulation is now being used as<br>3198 auxiliary training targets for GN1. The solid lines represent reconstructed<br>3199 information, whereas the dashed lines represent truth information. . . . .  | 83 |
| 3200 | 6.2 The inputs to GN1 are the two jet features ( $n_{\text{jf}} = 2$ ), and an array of $n_{\text{tracks}}$ ,<br>3201 where each track is described by 21 track features ( $n_{\text{tf}} = 21$ ) [2]. The jet<br>3202 features are copied for each of the tracks, and the combined jet-track vectors<br>3203 of length 23 form the inputs of GN1. . . . .  | 90 |

|      |   |    |
|------|---|----|
| 3204 | 6.3 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. . . . .   | 90 |
| 3210 | 6.4 Comparison between the DL1r and GN1 $b$ -tagging discriminant $D_b$ for jets in the $t\bar{t}$ sample [2]. The 85% WP and the 60% WP are marked by the solid (dashed) lines for GN1 (DL1r), representing respectively the loosest and tightest WPs typically used by analyses. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1. The distributions of the different jet flavours have been normalised to unity area.  | 95 |
| 3216 | 6.5 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 [2]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The $x$ -axis range is chosen to display the $b$ -jet tagging efficiencies usually probed in these regions of phase space. | 98 |
| 3223 | 6.6 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet tagging efficiency for jets in the $Z'$ sample with $250 < p_T < 5000$ GeV [2]. The ratio with respect to the performance of the DL1r algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. Binomial error bands are denoted by the shaded regions. The $x$ -axis range is chosen to display the $b$ -jet tagging efficiencies usually probed in these regions of phase space.     | 98 |
| 3230 | 6.7 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 [2]. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1 and GN1Lep. GN1 demonstrates improved performance with respect to DL1r accross the $p_T$ range shown. Binomial error bands are denoted by the shaded regions. . . . .   | 99 |

|      |  |     |
|------|--|-----|
| 3236 | 6.8 The $b$ -jet (left) and light-jet (right) rejections as a function of the $c$ -jet         |     |
| 3237 | tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$ [2]. The ratio to the |     |
| 3238 | performance of the DL1r algorithm is shown in the bottom panels. Binomial                      |     |
| 3239 | error bands are denoted by the shaded regions. The $x$ -axis range is chosen                   |     |
| 3240 | to display the $c$ -jet tagging efficiencies usually probed in these regions of                |     |
| 3241 | phase space. . . . .   | 101 |
| 3242 | 6.9 The $b$ -jet (left) and light-jet (right) rejections as a function of the $c$ -jet         |     |
| 3243 | tagging efficiency for $Z'$ jets with $250 < p_T < 5000 \text{ GeV}$ [2]. The ratio            |     |
| 3244 | to the performance of the DL1r algorithm is shown in the bottom panels.                        |     |
| 3245 | Binomial error bands are denoted by the shaded regions. The $x$ -axis range                    |     |
| 3246 | is chosen to display the $c$ -jet tagging efficiencies usually probed in these                 |     |
| 3247 | regions of phase space. . . . .  | 101 |
| 3248 | 6.10 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet        |     |
| 3249 | tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250 \text{ GeV}$ , for the nominal     |     |
| 3250 | GN1, in addition to configurations where no (GN1 No Aux), only the                             |     |
| 3251 | track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary                       |     |
| 3252 | objectives are deployed [2]. The ratio to the performance of the DL1r                          |     |
| 3253 | algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in                   |     |
| 3254 | the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1. . . . .                    | 103 |
| 3255 | 6.11 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet        |     |
| 3256 | tagging efficiency for $Z'$ jets with $250 < p_T < 5000 \text{ GeV}$ , for the nominal         |     |
| 3257 | GN1, in addition to configurations where no (GN1 No Aux), only the                             |     |
| 3258 | track classification (GN1 TC) or only the vertexing (GN1 Vert) auxiliary                       |     |
| 3259 | objectives are deployed [2]. The ratio to the performance of the DL1r                          |     |
| 3260 | algorithm is shown in the bottom panels. A value of $f_c = 0.018$ is used in                   |     |
| 3261 | the calculation of $D_b$ for DL1r and $f_c = 0.05$ is used for GN1. . . . .                    | 103 |
| 3262 | 6.12 A schematic showing the truth track origin and vertex information (left)                  |     |
| 3263 | compared with the predictions from GN1 (right). The diagrams show the                          |     |
| 3264 | truth (left) and predicted (right) structure of a $b$ -jet. The shaded black boxes             |     |
| 3265 | show the groupings of tracks into different vertices. Vertices reconstructed                   |     |
| 3266 | by SV1 and JetFitter are also marked. Class probabilities $p_b$ , $p_c$ and $p_u$ are          |     |
| 3267 | rounded to three significant figures. GN1 improves over SV1 and JetFitter                      |     |
| 3268 | by successfully determining the origins and vertex compatibility of all the                    |     |
| 3269 | tracks in the jet with the exception of the truth OtherSecondary track,                        |     |
| 3270 | which is misidentified as pileup. . . . .  | 105 |

|      |   |     |
|------|---|-----|
| 3271 | 6.13 A schematic showing the truth track origin and vertex information (left)                     |     |
| 3272 | compared with the predictions from GN1 (right). The diagrams show the                             |     |
| 3273 | truth (left) and predicted (right) structure of a $b$ -jet. The shaded black boxes                |     |
| 3274 | show the groupings of tracks into different vertices. Vertices reconstructed                      |     |
| 3275 | by SV1 and JetFitter are also marked. Class probabilities $p_b$ , $p_c$ and $p_u$ are             |     |
| 3276 | rounded to three significant figures. GN1 improves over SV1 and JetFitter                         |     |
| 3277 | by successfully determining the origins and vertex compatibility of all but                       |     |
| 3278 | one tracks in the jet. . . . .  | 105 |
| 3279 | 6.14 Heavy flavour vertex finding efficiency as a function of jet $p_T$ for $b$ -jets in          |     |
| 3280 | the $t\bar{t}$ sample using the exclusive (left) and inclusive (right) vertex finding             |     |
| 3281 | approaches. The exclusive vertexing approach includes single track vertices                       |     |
| 3282 | while the inclusive approach requires vertices to contain at least two tracks.                    |     |
| 3283 | Efficient vertex finding requires the recall of at least 65% of the tracks in                     |     |
| 3284 | the truth vertex, and allows no more than 50% of the tracks to be included                        |     |
| 3285 | incorrectly. Binomial error bands are denoted by the shaded regions. . .                          | 108 |
| 3286 | 6.15 Heavy flavour vertex finding efficiency as a function of jet $p_T$ for $b$ -jets in          |     |
| 3287 | the $Z'$ sample using the exclusive (left) and inclusive (right) vertex finding                   |     |
| 3288 | approaches. The exclusive vertexing approach includes single track vertices                       |     |
| 3289 | while the inclusive approach requires vertices to contain at least two tracks.                    |     |
| 3290 | Efficient vertex finding requires the recall of at least 65% of the tracks in                     |     |
| 3291 | the truth vertex, and allows no more than 50% of the tracks to be included                        |     |
| 3292 | incorrectly. Binomial error bands are denoted by the shaded regions. . .                          | 108 |
| 3293 | 6.16 Vertex finding efficiency as a function of jet $p_T$ for light-jets in the $t\bar{t}$ sample |     |
| 3294 | using the exclusive (left) and inclusive (right) vertex finding approaches.                       |     |
| 3295 | The exclusive vertexing approach includes single track vertices while the                         |     |
| 3296 | inclusive approach requires vertices to contain at least two tracks. Efficient                    |     |
| 3297 | vertex finding requires the recall of at least 65% of the tracks in the truth                     |     |
| 3298 | vertex, and allows no more than 50% of the tracks to be included incorrectly.                     |     |
| 3299 | Binomial error bands are denoted by the shaded regions. . . . .                                   | 109 |

|      |  |     |
|------|--|-----|
| 3300 | 6.17 Vertex finding efficiency as a function of jet $p_T$ for light-jets in the $Z'$ sample using the exclusive (left) and inclusive (right) vertex finding approaches. The exclusive vertexing approach includes single track vertices while the inclusive approach requires vertices to contain at least two tracks. Efficient vertex finding requires the recall of at least 65% of the tracks in the truth vertex, and allows no more than 50% of the tracks to be included incorrectly. Binomial error bands are denoted by the shaded regions. . . . . | 110 |
| 3301 |  |     |
| 3302 |  |     |
| 3303 |  |     |
| 3304 |  |     |
| 3305 |  |     |
| 3306 |  |     |
| 3307 | 6.18 ROC curves for the different groups of truth origin labels defined in Table 5.1 for jets in the $t\bar{t}$ sample (left) and jets in the $Z'$ sample (right) [2]. 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. . . . .   | 112 |
| 3308 |  |     |
| 3309 |  |     |
| 3310 |  |     |
| 3311 |  |     |
| 3312 |  |     |
| 3313 |  |     |
| 3314 | 6.19 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet tagging efficiency for $t\bar{t}$ jets with $20 < p_T < 250$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1d and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. . . . .   | 114 |
| 3315 |  |     |
| 3316 |  |     |
| 3317 |  |     |
| 3318 |  |     |
| 3319 |  |     |
| 3320 | 6.20 The $c$ -jet (left) and light-jet (right) rejections as a function of the $b$ -jet tagging efficiency for $Z'$ jets with $250 < p_T < 5000$ GeV, for the looser track selection trainings of GN1. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. A value of $f_c = 0.018$ is used in the calculation of $D_b$ for DL1d and $f_c = 0.05$ is used for GN1. Binomial error bands are denoted by the shaded regions. . . . .   | 114 |
| 3321 |  |     |
| 3322 |  |     |
| 3323 |  |     |
| 3324 |  |     |
| 3325 |  |     |
| 3326 | 6.21 The light-jet rejection as a function of the $b$ -jet efficiency jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of mass energy $\sqrt{s} = 13.6$ TeV [125]. The ratio to the performance of the DL1d algorithm [124] is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for b-tagging. . . . .   | 115 |
| 3327 |  |     |
| 3328 |  |     |
| 3329 |  |     |
| 3330 |  |     |
| 3331 |  |     |

---

|      |  |     |
|------|--|-----|
| 3332 | 6.22 The $c$ -jet rejection (left) and light-jet rejection (right) as a function of the $b$ -jet efficiency for jets in the $t\bar{t}$ sample with $20 < p_T < 250$ GeV for events with a centre of mass energy $\sqrt{s} = 14$ TeV [126]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. The purple vertical dashed lines represent the most common working points used for b-tagging. . . . . | 116 |
| 3338 | 6.23 The $c$ -jet rejection (left) and light-jet rejection (right) as a function of the $b$ -jet efficiency for jets in the $Z'$ sample with $250 < p_T < 5000$ GeV for events with a centre of mass energy $\sqrt{s} = 14$ TeV [126]. The ratio to the performance of the DL1d algorithm is shown in the bottom panels. Binomial error bands are denoted by the shaded regions. . . . .   | 116 |
| 3343 | 7.1 Diagrams of the signal process (top) and examples of the the 0-lepton and 1-lepton $t\bar{t}$ backgrounds (middle, bottom). Object to the right of centre are reconstructed within the large- $R$ jet. For the backgrounds, the large- $R$ jet contains a mis-tagged $c$ - or light-jet. Not shown is the 0-lepton background resulting from $W \rightarrow \ell\nu$ , where the lepton is out of acceptance or not reconstructed. . . . .   | 127 |
| 3349 | 7.2 Leading large- $R$ jet mass for the $Z$ and $W+hf$ processes in the HP SR of the 0-lepton channel [174]. The renormalisation scale $\mu_r$ has been varied by a factor of 2 (1up) and 0.5 (1down). An exponential function is fitted to the ratio between the nominal and alternate samples. . . . .   | 137 |
| 3353 | 7.3 Leading large- $R$ jet $m_J$ inclusive in $p_T^V$ for the $V+hf$ process modelled using both the SHERPA and MADGRAPH samples [174]. The Kolmogorov-Smirnov test and $\chi^2/ndf$ are shown on the plots. . . . .   | 138 |
| 3356 | 7.4 Leading large- $R$ jet mass distribution for the $WZ$ and $ZZ$ processes, inclusive across all signal regions and lepton channels [174]. The renormalisation scale $\mu_R$ has been varied by a factor of 2 (1up) and 0.5 (1down). The red line shape shows the fitting results of the hyperbolic tangent function. . . . .  | 140 |
| 3360 | 7.5 The comparison on $m_J$ shapes between SHERPA and POWHEG+PYTHIA 8 samples from $WZ$ and $ZZ$ process in high and low purity signal regions, integrated over $p_T^V$ regions and the 0- and 1-lepton channels [174]. The dashed green line shows the fitted third order polynomial function and the blue lines show the function after a protection is added in the high mass region. . . . .   | 141 |

---

|      |   |     |
|------|---|-----|
| 3366 | 7.6 The $m_J$ post-fit distributions in (top) the 0-, (middle) 1- and (bottom) 2-lepton SRs for (left) $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (right) $p_T^V \geq 400 \text{ GeV}$ . The LP and HP regions are merged for the 0-lepton and 1-lepton channels. The fitted background contributions are shown as filled histograms. The Higgs boson signal ( $m_H = 125 \text{ GeV}$ ) is shown as a filled histogram and is normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and as an unstacked unfilled histogram, scaled by the SM prediction times a factor of two. The size of the combined on the sum of the fitted signal and background is shown in the hatched band. The highest bin contains the overflow [129]. . . . .  | 147 |
| 3367 |   |     |
| 3368 |   |     |
| 3369 |   |     |
| 3370 |   |     |
| 3371 |   |     |
| 3372 |   |     |
| 3373 |   |     |
| 3374 |   |     |
| 3375 |   |     |
| 3376 | 7.7 The $m_J$ post-fit distributions in the $t\bar{t}$ control region for (top) the 0-lepton channel and the 1-lepton channel for $250 \text{ GeV} < p_T^V < 400 \text{ GeV}$ and (bottom) the 0-lepton channel and the 1-lepton channel for $p_T^V > 400 \text{ GeV}$ . The background contributions after the likelihood fit are shown as filled histograms. The Higgs boson signal ( $m_h = 125 \text{ GeV}$ ) is shown as a filled histogram on top of the fitted backgrounds normalised to the signal yield extracted from data ( $\mu_{VH}^{bb} = 0.72$ ), and unstacked as an unfilled histogram, scaled by the SM prediction times a factor of 2. The size of the combined statistical and systematic uncertainty for the sum of the fitted signal and background is indicated by the hatched band. The highest bin in the distributions contains the overflow [129]. . . . .   | 148 |
| 3377 |   |     |
| 3378 |   |     |
| 3379 |   |     |
| 3380 |   |     |
| 3381 |   |     |
| 3382 |   |     |
| 3383 |   |     |
| 3384 |   |     |
| 3385 |   |     |
| 3386 |   |     |
| 3387 | 7.8 $m_J$ distribution in data after subtraction of all backgrounds except for the $WZ$ and $ZZ$ diboson processes. The contributions from all lepton channels and signal regions are summed and weighted by their respective values of the ratio of fitted Higgs boson signal and background yields. The expected contribution of the associated $WH$ and $ZH$ production of a SM Higgs boson with $m_H = 125 \text{ GeV}$ is shown scaled by the measured combined signal strength ( $\mu_{VH}^{bb} = 0.72$ ). The diboson contribution is normalised to its best-fit value of $\mu_{VZ}^{bb} = 0.91$ . The size of the combined statistical and systematic uncertainty is indicated by the hatched band. This error band is computed from a full signal-plus-background fit including all the systematic uncertainties defined in Section 7.2, except for the $VH/VZ$ experimental and theory uncertainties [129]. . . . . | 149 |
| 3388 |   |     |
| 3389 |   |     |
| 3390 |   |     |
| 3391 |   |     |
| 3392 |   |     |
| 3393 |   |     |
| 3394 |   |     |
| 3395 |   |     |
| 3396 |   |     |
| 3397 |   |     |
| 3398 |   |     |

|      |  |     |
|------|--|-----|
| 3399 | 7.9 Signal strength compatibility test between the (3+1)-POI fit (with the three                               |     |
| 3400 | lepton channels splitted) and the default (1+1)-POI fit. The compatibility                                     |     |
| 3401 | of the three channels is evaluated via a $\chi^2$ difference test and results in a                             |     |
| 3402 | p-value of 49% [129]. . . . .  | 150 |
| 3403 | 7.10 Impact of systematic uncertainties on the fitted $VH$ signal-strength pa-                                 |     |
| 3404 | rameter $\mu_{VH}^{bb}$ sorted in decreasing order. The boxes show the variations                              |     |
| 3405 | of $\hat{\mu}$ , referring to the top $x$ -axis, when fixing the corresponding individual                      |     |
| 3406 | nuisance parameter to its post-fit value modified upwards or downwards by                                      |     |
| 3407 | its post-fit uncertainty, i.e. $\hat{\theta} \pm \sigma_{\hat{\theta}}$ , and repeating the fit. The impact of |     |
| 3408 | up- and down-variations can be distinguished via the dashed and plane box                                      |     |
| 3409 | fillings. The yellow boxes show the pre-fit impact (top $x$ -axis) by varying                                  |     |
| 3410 | each nuisance parameter by $\pm 1$ . The filled circles show the deviation of the                              |     |
| 3411 | fitted value for each nuisance parameter, $\hat{\theta}$ , from their nominal input value                      |     |
| 3412 | $\theta_0$ expressed in standard deviations with respect to their nominal uncer-                               |     |
| 3413 | tainties $\Delta\theta$ (bottom $x$ -axis). The error bars show the post-fit uncertainties                     |     |
| 3414 | on $\hat{\theta}$ with respect to their nominal uncertainties. The open circles show                           |     |
| 3415 | the fitted values and uncertainties of the normalization parameters that are                                   |     |
| 3416 | freely floating in the fit. Pre-fit, these parameters have a value of one [174].                               | 152 |

# 3417 List of tables

|                     |  |    |
|---------------------|--|----|
| <small>3418</small> | 2.1 The fermions of the SM [15]. Three generations of particles are present.<br><small>3419</small> Also present (unlisted) are the antiparticles, which are identical to the<br><small>3420</small> particles up to a reversed charge sign. . . . .   | 15 |
| <small>3421</small> | 2.2 The bosons of the SM [15]. The photon, weak bosons and gluons are gauge<br><small>3422</small> bosons arising from gauge symmetries, and carry the four fundamental forces<br><small>3423</small> of the SM. The recently discovered Higgs boson is the only fundamental<br><small>3424</small> scalar particle in the SM. . . . .   | 16 |
| <small>3425</small> | 3.1 Overview of the different LHC runs [33, 34]. The average number of interac-<br><small>3426</small> tions per bunch-crossing is denoted as $\langle \mu \rangle$ (see Section 3.2.3), and is here<br><small>3427</small> averaged over the entire run. Numbers for Run 3 are preliminary and are<br><small>3428</small> only provided to give an indication of expected performance. . . . .  | 29 |
| <small>3429</small> | 5.1 Truth origins which are used to categorise the physics process that led<br><small>3430</small> to the production of a track. Tracks are matched to charged particles<br><small>3431</small> using the truth-matching probability [56]. A truth-matching probability<br><small>3432</small> of less than 0.5 indicates that reconstructed track parameters are likely<br><small>3433</small> to be mismeasured and may not correspond to the trajectory of a single<br><small>3434</small> charged particle. The “OtherSecondary” origin includes tracks from photon<br><small>3435</small> conversions, $K_S^0$ and $\Lambda^0$ decays, and hadronic interactions. . . . . | 72 |
| <small>3436</small> | 5.2 Input features to the fake track classification NN. Basic jet kinematics, along<br><small>3437</small> with information about the reconstructed track parameters and constituent<br><small>3438</small> hits are used. Shared hits, are hits used on multiple tracks which have not<br><small>3439</small> been classified as split by the cluster-splitting neural networks [56], while<br><small>3440</small> split hits are hits used by multiple tracks which have been identified as<br><small>3441</small> merged, and therefore split. . . . .  | 75 |

|      |   |     |
|------|---|-----|
| 3442 | 5.3 Quality selections applied to tracks, where $d_0$ is the transverse IP of the track, $z_0$ is the longitudinal IP with respect to the PV and $\theta$ is the track polar angle (see Section 3.2.2 for the IP definitions). Shared hits are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [56]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. . . . .  | 76  |
| 3448 | 5.4 Hyperparameter for the track classification model . . . . .   | 77  |
| 3449 | 5.5 Good and fake track selection efficiencies for the combined $t\bar{t}$ and $Z'$ samples. Two working points are defined, cutting on the NN output at 0.06 and 0.12. The continuous output of the model allows for the tuning of good and fake track identification efficiencies. . . . .  | 78  |
| 3453 | 5.6 Cut values for the fake and $b$ -hadron decay track MVAs for the two defined working points. Working point “B” cuts more aggressively on the MVA outputs than WP “A”, removing more fake tracks but resulting in an increased loss of signal tracks (which here are all $b$ -hadron decay tracks). . . . .  | 79  |
| 3457 | 6.1 Input features to the GN1 model [2]. Basic jet kinematics, along with information about the reconstructed track parameters and constituent hits are used. Shared hits, are hits used on multiple tracks which have not been classified as split by the cluster-splitting neural networks [56], while split hits are hits used on multiple tracks which have been identified as merged. A hole is a missing hit, where one is expected, on a layer between two other hits on a track. The track leptonID is an additional input to the GN1Lep model. . . . . | 87  |
| 3465 | 6.2 A summary of GN1’s different classification networks used for the various training objectives, adapted from Ref. [2]. The hidden layers column contains a list specifying the number of neurons in each layer. . . . .  | 92  |
| 3468 | 6.3 The area under the ROC curves (AUC) for the track classification from GN1, compared to a standard multilayer perceptron (MLP) trained on a per-track basis. The unweighted mean AUC over the origin classes and weighted mean AUC (using as a weight the fraction of tracks from the given origin) is provided. GN1, which uses an architecture that allows track origins to be classified in a conditional manner as discussed in Section 6.3.3, outperforms the MLP model for both $t\bar{t}$ and $Z'$ jets. . . . .                                      | 111 |

|      |   |     |
|------|---|-----|
| 3475 | 7.1 Summary of the definitions of the different analysis regions [129]. Signal<br>enriched regions are marked with the label SR. There are regions with rela-<br>tively large signal purity (HP SR) and with low purity (LP SR). Background<br>enriched regions are marked with the label CR. The shorthand “add” stands<br>for additional small- $R$ jets, i.e. number of small- $R$ jets not matched to the<br>Higgs-jet candidate. The medium and high $p_T^V$ regions are referred to as<br>$Mp_T^V$ and $Hp_T^V$ , respectively. . . . .   | 121 |
| 3482 | 7.2 Signal and background processes with the corresponding generators used for the<br>nominal samples [129]. If not specified, the order of the cross-section calculation<br>refers to the expansion in the strong coupling constant ( $\alpha_s$ ). (★) The events<br>were generated using the first PDF in the NNPDF3.0NLO set and subsequently<br>reweighted to the PDF4LHC15NLO set [163] using the internal algorithm in<br>POWHEG-Box v2. (†) The NNLO(QCD)+NLO(EW) cross-section calculation for<br>the $pp \rightarrow ZH$ process already includes the $gg \rightarrow ZH$ contribution. The $qq \rightarrow ZH$<br>process is normalised using the cross-section for the $pp \rightarrow ZH$ process, after<br>subtracting the $gg \rightarrow ZH$ contribution. An additional scale factor is applied to<br>the $qq \rightarrow VH$ processes as a function of the transverse momentum of the vector<br>boson, to account for electroweak (EW) corrections at NLO. This makes use of<br>the $VH$ differential cross-section computed with HAWK [164, 165]. . . . . | 123 |
| 3494 | 7.3 Selections applied to baseline and signal electrons and muons. . . . .  | 124 |
| 3495 | 7.4 Event selection requirements for the boosted $VH$ , $H \rightarrow b\bar{b}$ analysis channels<br>and sub-channels [129]. Various channel dependent selections are used to<br>maximise the signal efficiency and reduce the number of background events<br>in each analysis region. The $\min[\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{small-}R \text{ jets})]$ selection is used to<br>remove jets when the missing transverse momentum $\mathbf{E}_T^{\text{miss}}$ is pointing in the<br>direction of the Higgs candidate, and the $\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{E}_{T, \text{trk}})$ is used to reject<br>events where the calorimeter missing transverse momentum $\mathbf{E}_T^{\text{miss}}$ is not<br>pointing in the direction of the track-based missing transverse momentum<br>$\mathbf{E}_{T, \text{trk}}^{\text{miss}}$ . The $\Delta y(V, H_{\text{cand}})$ quantifies the rapidity difference between the<br>reconstructed vector boson and Higgs candidate. . . . .   | 126 |
| 3505 | 7.5 Different sources of uncertainty (i.e. variations in the model) considered for<br>the $V + \text{jets}$ background, and the corresponding implementation. For each<br>uncertainty, acceptance and shape uncertainties are derived. . . . .  | 130 |

|      |   |     |
|------|---|-----|
| 3508 | 7.6 0-lepton $W$ +jets nominal sample flavour composition and total event yield [174].<br>3509 $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see<br>3510      Table 7.1). . . . .   | 133 |
| 3511 | 7.7 1-lepton $W$ +jets nominal sample flavour composition and total event yield [174].<br>3512 $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see<br>3513      Table 7.1). . . . .   | 133 |
| 3514 | 7.8 0-lepton $Z$ +jets nominal sample flavour composition and total event yield [174].<br>3515 $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see<br>3516      Table 7.1). . . . .   | 133 |
| 3517 | 7.9 2-lepton $Z$ +jets nominal sample flavour composition and total event yield [174].<br>3518 $Mp_T^V$ refers to the medium $p_T^V$ region, and $Hp_T^V$ refers to the high $p_T^V$ region (see<br>3519      Table 7.1). . . . .   | 134 |
| 3520 | 7.10 $V$ +jets acceptance uncertainties [174]. $W$ +jets SR and CR uncertainties<br>3521      marked with a superscript † are correlated. The 1L $W$ +jets H/M uncertainty<br>3522      marked by * is applied as independent and uncorrelated NPs in both HP<br>3523      and LP signal regions. . . . .   | 135 |
| 3524 | 7.11 Diboson acceptance uncertainties [174]. All uncertainties except channel ex-<br>3525      trapolation uncertainties are fully correlated between $ZZ$ and $WZ$ processes<br>3526      and channels. . . . .  | 139 |
| 3527 | 7.12 Factors applied to the nominal normalisations of the $t\bar{t}$ , $W$ +hf, and $Z$ +hf<br>3528      backgrounds, as obtained from the likelihood fit [129]. The errors represent<br>3529      the combined statistical and systematic uncertainties. . . . .   | 145 |
| 3530 | 7.13 Breakdown of the absolute contributions to the uncertainty on the signal<br>3531      strength $\mu_{VH}^{bb}$ obtained from the (1+1)-POI fit. The average impact repre-<br>3532      sents the average between the positive and negative uncertainties on $\mu_{VH}^{bb}$ .<br>3533      The sum in quadrature of the systematic uncertainties attached to the cate-<br>3534      gories differs from the total systematic uncertainty due to correlations [174].<br>3535      . . . . . | 153 |